

The large-scale correlations of multi-cell densities and profiles: implications for cosmic variance estimates

Sandrine Codis^{1,2*}, Francis Bernardeau^{2,3}, Christophe Pichon^{2,4}

¹ *Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON M5S 3H8, Canada*

² *Sorbonne Universités, UPMC Univ Paris 6 & CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France*

³ *CNRS & CEA, UMR 3681, Institut de Physique Théorique, F-91191 Gif-sur-Yvette, France*

⁴ *Korea Institute of Advanced Studies (KIAS) 85 Hoegiro, Dongdaemun-gu, Seoul, 02455, Republic of Korea*

May 9, 2016

ABSTRACT

In order to quantify the error budget in the measured probability distribution functions of cell densities, the two-point statistics of cosmic densities in concentric spheres is investigated. Bias functions are introduced as the ratio of their two-point correlation function to the two-point correlation of the underlying dark matter distribution. They describe how cell densities are spatially correlated. They are computed here via the so-called large deviation principle in the quasi-linear regime. Their *large-separation limit* is presented and successfully compared to simulations for density and density slopes: this regime is shown to be rapidly reached allowing to get sub-percent precision for a wide range of densities and variances. The corresponding asymptotic limit provides an estimate of the cosmic variance of standard concentric cell statistics applied to finite surveys. More generally, no assumption on the separation is required for some specific moments of the two-point statistics, for instance when predicting the generating function of cumulants containing any powers of concentric densities in one location and one power of density at some *arbitrary* distance from the rest. This exact “one external leg” cumulant generating function is used in particular to probe the rate of convergence of the large-separation approximation.

Key words: cosmology: theory — large-scale structure of Universe — methods: analytical, numerical

1 INTRODUCTION

The geometry of the large-scale structure of the Universe puts very tight constraints on cosmological models. Deep spectroscopic surveys, like Euclid (Laureijs et al. 2011) or DESI (Levi et al. 2013), will soon allow us to study the details of structure formation at different epochs with unrivaled precision and therefore offer insight into the engine of cosmic acceleration. In order to reach percent precision on the equation of state of dark energy, astronomers are facing various challenges: non-linear gravitational evolution (Bernardeau et al. 2002), redshift space distortions (Kaiser 1987; Taruya et al. 2010), bias (Kaiser 1984; Dekel & Rees 1987), intrinsic alignments (Kiessling et al. 2015), baryonic physics (Schneider & Teyssier 2015) to name a few. Probing the non-linear regime increases the number of modes used to better constrain cosmological parameters. Hence, theorists need to investigate alternatives to the standard N-point correlation functions (Scoccimarro et al. 1998) that Perturba-

tion Theory can only predict in the weakly non-linear regime (Lazanu et al. 2015). They must find new observables that can be predicted from first principles, and do not rely solely on very large simulations of the Universe produced with hundreds of millions of CPU hours.

It has been argued (Bernardeau et al. 2014, 2015; Uhlemann et al. 2015) that the statistics of cosmic densities in concentric spheres can leverage cosmic parameters competitively, as the corresponding spherical symmetry allows for analytical predictions in the mildly non-linear regime, beyond what is commonly achievable via other statistics such as correlation functions. Indeed, the zero variance limit of the cumulant generating functions yields estimates of the joint probability distribution function (PDF hereafter) which seems to match simulations in the regime of variances of order unity as shown by Valageas (2002); Bernardeau et al. (2014, 2015) building upon some earlier investigations by Balian & Schaeffer (1989); Bernardeau (1992). This success has been recently shown to correspond to a large-deviation principle in the context of cosmic structure formation (Bernardeau & Reimberg 2015) which

* codis@cita.utoronto.ca

is based on the sole assumption that the variance of the field is small. The corresponding predictions for the successive reduced cumulants were shown to be in excellent agreement with simulations for scale above a few Mpc/h (e.g. [Baugh et al. \(1995\)](#)). It has to be contrasted with the commonly used polyspectra predicted by Perturbation Theory which typically require the fluctuations of the field to be small everywhere. For instance, unless large-deviation predictions, the Edgeworth expansion typically breaks down for $|\delta| > \sigma$.

Hence it is of interest to quantify the error budget for such large-deviation estimators, while accounting for the expected long-range clustering within realistic surveys of finite extent. Indeed, in practice, measurements of cosmic densities cannot be carried out for different realisations i.e. different universes but in one finite part of our Universe as mapped by surveys like the SDSS ([Abazajian et al. 2003](#)) or DES ([Sánchez et al. 2014](#)). The density in spheres drawn from those surveys are not independent. Consequently, this dependence induces errors which, at first order, are dominated by the two-point correlation between spheres, as shown by [Maurogordato & Lachieze-Rey \(1987\)](#); [Colombi et al. \(1995\)](#); [Szapudi & Colombi \(1996\)](#); [Moster et al. \(2011\)](#). Estimating the two-point correlations of concentric cosmic densities ([Bernardeau 1996](#); [Szapudi et al. 1999](#)) is therefore important to mitigate the cosmic variance on the measurement of their one-point statistics. Once a model for these correlations exists, it can be integrated into maximum likelihood estimators for the underlying cosmic parameters.

Besides error statistics, the study of this two-point clustering statistics of concentric spheres is also interesting in its own right, as it allows one to investigate how the densest regions of space – where dark halos usually reside – are clustered in the quasi-linear regime, which in turn sheds light on the so-called biasing between dark matter and halos: a common, apparently good assumption ([Ludlow & Porciani 2011](#)), is that haloes correspond to peaks of the density field, they are therefore not a fair tracer of that field. For Gaussian random fields, [Kaiser \(1984\)](#) showed that in the high contrast ν , high separation limit, the correlation function, $\xi_{>\nu}$, between two regions lying above a threshold ν reads

$$\xi_{>\nu} \approx \nu^2 \xi, \quad (1)$$

so that the correlation function of high-density regions decreases more slowly than the density field correlation function, ξ , with an amplification factor or clustering bias that is proportional to the threshold squared. This analysis can also be restricted to the peaks of the density field above a given threshold following the seminal paper by [Bardeen et al. \(1986\)](#).

Building upon the idea that the density of large clusters must be strongly clustered compared to the density of galaxies, [Davis et al. \(1985\)](#) popularized the notion of linear bias between the galaxy distribution and dark matter density field

$$\delta_g = b_1 \delta_{\text{DM}}, \quad (2)$$

where b_1 is assumed to be a constant which was shown to be a good approximation on sufficiently large scales ([Manera & Gaztañaga 2011](#)). However, given the complexity of galaxy formation, the validity of this approximation

is likely to be somewhat narrow. [Fry & Gaztañaga \(1993\)](#) proposed to extend the linear bias approach and to introduce the so-called local bias model in which the full Taylor expansion of the relation $\delta_g = \mathcal{F}(\delta_{\text{DM}})$ is considered. This parametrization can be studied from a Eulerian or a Lagrangian point of view, meaning at final or initial time, ([Catelan et al. 2000](#)) and can account for stochasticity ([Dekel & Lahav \(1999\)](#)), time-evolution ([Nusser & Davis \(1994\)](#); [Fry \(1996\)](#); [Tegmark & Peebles \(1998\)](#)), non-linearity ([Pen \(1998\)](#); [Guo & Jing \(2009\)](#)), scale-dependence ([Lumsden et al. \(1989\)](#); [Giannantonio & Porciani \(2010\)](#)) or non-locality in time ([Senatore 2015](#)). Non-local effects can also be addressed by parametrizing the dependence of the galaxy density with other operators such as the tidal tensor or velocity shear as long as they preserve the symmetry and equivalence ([McDonald & Roy 2009](#); [Baldauf et al. 2012](#); [Mirbabayi et al. 2015](#)). De facto, this bias is very likely to depend on the population of galaxies and to be not only a function of the density but also temperature, merging history and other galaxy properties, introducing some scatter in the galaxy-matter density relation. Quantifying bias is crucial in cosmology if one wants to measure the cosmological parameters encoded in ξ from, e.g. the correlation function of galaxies. It is also an interesting quantity to measure in its own right, as it carries information on the physics of galaxy formation. Numerical simulations are very efficient at accurately predicting and calibrating galactic observables. Improvements of our understanding of galaxy formation and its implementation in simulations is a timely and fast moving topic of research.

However, it has to be noted that the scale-dependence of the halo bias (defined as the ratio of the halo-halo to dark matter correlation functions) is strong on small scales but remains weak with typical deviations of less than a few percents on scales above $20 \text{ Mpc}/h$ ([Crocce et al. 2015](#)). It would therefore be interesting to predict the bias of specific regions (for instance high density and/or negative slope as a proxy for dark halos) from first principles not only in the linear stage of structure formation but also in the subsequent mildly non-linear regime.

The aim of this paper is to quantify the effect of cosmic variance on count-in-cells statistics in the mildly non-linear regime. Specifically we will estimate the two-point correlations of concentric cosmic densities at different positions in the field in the (not so) large separation limit, while relying on the symmetric framework of the spherical collapse, which leads to surprisingly accurate predictions. It will allow us in particular to define the bias associated with concentric cosmic densities and provide accurate predictions up to a few Mpc/h .

This paper is organized as follows. Section 2 predicts the bias and error budget expected in standard concentric count-in-cells. Section 3 shows how to compute the generating function of the cumulants containing any power of the densities in one location and one power of the density at an arbitrary distance. Section 4 predicts the bias functions in the large separation limit while section 5 applies the formalism to realistic power spectra. Section 6 validates this large-separation and large-deviation approximation on simulations before Section 7 illustrates how bias functions can be used to predict the expected error budget when estimating the PDF of the cosmic density in concentric spheres. Finally,

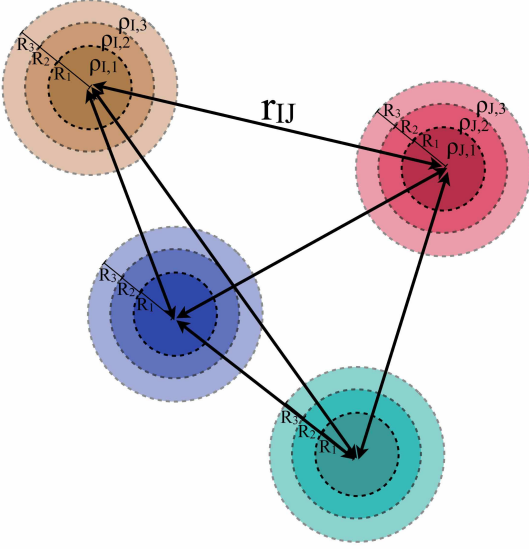


Figure 1. The configuration of spherical cells considered in this paper which is made of multiple sets of concentric spheres separated by distances r_{IJ} . Their respective density, $\rho_{I,i}$, corresponds to a set of n spheres of same radii $R_{I,i} \equiv R_i$.

Section 8 wraps up. Appendices A, B, C, D, E and F respectively present the radii decimation used in the main text, an argument in favour of the large-distance factorizations, some calculations of cosmic variance for discrete counts, a consistency check of the measurements of the bias function by two distinct methods and analytical asymptotes for the density and slope bias.

2 STATISTICS OF SETS OF COUNT-IN-CELLS

In a survey or a simulation, measurements of concentric densities are carried as follows: sets of concentric cells are drawn randomly or regularly across the field at some finite distance to each other. In cosmology, this field has long-range correlations which will break the assumption that each set of concentric spheres can be considered independently. In order to estimate the induced bias and variance, let us study the joint statistics of these sets.

2.1 The bias function of concentric spheres

Let us consider multiple sets (labelled from $I = 1$ to N_t) of n concentric spheres (labelled from $k = 1$ to n) of radii $R_{I,k} \equiv R_k$ separated by distances r_{IJ} , and define the corresponding measured densities $\{\rho_{I,k}\}$ (see Fig. 1 for an illustration). The joint PDF of those N_t sets,

$$\mathcal{P}(\{\rho_{1,k}\}, \dots, \{\rho_{N_t,k}\}; \{r_{IJ}\}), \quad (3)$$

determines the full hierarchy of cumulants of the form

$$\langle \rho_{1,1}^{p_1} \dots \rho_{1,n}^{p_n} \rho_{2,1}^{q_1} \dots \rho_{2,n}^{q_n} \dots \rho_{N_t,1}^{s_1} \dots \rho_{N_t,n}^{s_n} \rangle_c. \quad (4)$$

The purpose of this paper is to estimate the joint PDF $\mathcal{P}(\{\rho_{1,k}\}, \dots, \{\rho_{N_t,k}\}; \{r_{IJ}\})$ in the large-separation limit,

where $r_{IJ} \gg R_{\max} = \max_j R_j$. In this limit, we will demonstrate in Section 4.2 that this PDF reads

$$\mathcal{P}(\{\rho_{1,k}\}, \dots, \{\rho_{N_t,k}\}; \{r_{IJ}\} \gg R_j) = \prod_{I=1}^{N_t} \mathcal{P}(\{\rho_{I,k}\}) \left[1 + \sum_{I < J} b(\{\rho_{I,k}\}) b(\{\rho_{J,k}\}) \xi(r_{IJ}) \right], \quad (5)$$

where $\Pi_I \mathcal{P}(\{\rho_{I,k}\})$ is the product of one-point PDFs, $\xi(r)$ is the underlying dark matter correlation function, and $b(\{\rho_{I,k}\})$ is some local bias function for the set I of n concentric spheres. This is the count-in-cell analog of the so-called peak-background-split or clustering bias. Equation (5) is the key result of this paper and will be used in the following sections to compute $b(\{\rho_{I,k}\})$ whose final expression is given by equations (36) and (42) below. We will also show in Section 4.4 that the bias obeys $\int b(\rho) \mathcal{P}(\rho) d\rho = 0$ and $\int \rho b(\rho) \mathcal{P}(\rho) d\rho = 1$ so that the N -point PDF given in equation (5) is normalised and its marginal in one location is exactly given by the one-point PDF.

Equation (5) allows us to define the excess probability of having the sets of densities $\{\rho_{1,k}\}, \dots, \{\rho_{N_t,k}\}$ separated by $\{r_{IJ}\}$ as

$$\xi_{N_t}(\{\rho_{1,k}\}, \dots, \{\rho_{N_t,k}\}) = \sum_{I < J} b(\{\rho_{I,k}\}) b(\{\rho_{J,k}\}) \xi(r_{IJ}). \quad (6)$$

From equation (6), we see that the error in assuming that the draws of concentric densities in simulations are independent scales like the dark matter correlation¹.

2.2 The bias and variance of concentric cumulants

Let us now define the arithmetic mean over sets of concentric spheres as

$$\overline{\rho_1^{p_1} \dots \rho_n^{p_n}} \equiv \frac{1}{N_t} \sum_I \rho_{I,1}^{p_1} \dots \rho_{I,n}^{p_n}. \quad (7)$$

This quantity naturally corresponds to what astronomers would measure in practice (spatial averages rather than ensemble averages). Our purpose is to quantify the bias and the expected cosmic variance of this estimator. Given equation (5), one can check that the expectation of the arithmetic estimator defined by equation (7) obeys

$$\overline{\langle \rho_1^{p_1} \dots \rho_n^{p_n} \rangle_c} = \langle \rho_1^{p_1} \dots \rho_n^{p_n} \rangle_c, \quad (8)$$

so that the mean of the estimator given by equation (7) is unbiased at large distances.

Let us now estimate the cross correlation of this estimator, $C_{\mathbf{pq}} \equiv \overline{\langle \rho_1^{p_1} \dots \rho_n^{p_n} \rho_1^{q_1} \dots \rho_n^{q_n} \rangle_c}$ and express it in terms of moments of the bias function

$$C_{\mathbf{pq}} = \frac{1}{N_t} \langle \rho_1^{p_1+q_1} \dots \rho_n^{p_n+q_n} \rangle_c + \frac{1}{N_t^2} \sum_{I \neq J} \xi(r_{IJ}) \times \langle b(\rho_1 \dots \rho_n) \rho_1^{p_1} \dots \rho_n^{p_n} \rangle_c \langle b(\rho_1 \dots \rho_n) \rho_1^{q_1} \dots \rho_n^{q_n} \rangle_c. \quad (9)$$

The first term in equation (9) is the error on the mean which is the typical error if draws are independent. The correlations between the draws – i.e the cells – lead to an additional source of errors encoded in the second term which

¹ In analogy with the corresponding situation for peaks, we can anticipate corrections involving derivative of the dark matter correlation at shorter separations.

corresponds to the bias function. Note that as expected, in the very large separation limit where $\xi(r_{\text{IJ}}) \rightarrow \delta_{\text{IJ}}$, we get $C_{\mathbf{pq}} \rightarrow C_{\mathbf{pq}}^0/N_t$ where $C_{\mathbf{pq}}^0 = \langle \rho_1^{p_1+q_1} \dots \rho_n^{p_n+q_n} \rangle_c$ and δ_{IJ} is the Kronecker delta function here.

2.3 Errors on the PDF

Let us finally quantify the cosmic variance on the estimate of the one-cell PDF when measuring densities in a finite number N_t of spheres. In this case, it is necessary to take into account the discreteness of the counts and the size of the bins of density.

One can show (see Appendix C for details) that in the Poisson limit, the number N of spheres with density in the interval $\Delta = [\rho - \Delta\rho/2, \rho + \Delta\rho/2]$ is unbiased ($\langle N \rangle = \bar{N}$) and has variance

$$\langle N^2 \rangle - \langle N \rangle^2 = \bar{N} + b^2 \xi \bar{N}^2, \quad (10)$$

where ξ is the mean correlation between the spheres, $\xi = \sum_{i \neq j} \xi(r_{\text{IJ}})/[N_t(N_t - 1)]$, and $\bar{N} = pN_t$ with $p = \int_{\Delta} d\rho \mathcal{P}(\rho)$ is the expected number of spheres with density in the interval considered. For a large enough number of spheres, sampling errors can therefore be neglected and the cosmic variance is directly proportional to b^2 where here b is defined as the mean density bias in the bin

$$b = \int_{\Delta} \mathcal{P}(\rho) b(\rho) d\rho / \int_{\Delta} \mathcal{P}(\rho) d\rho, \quad (11)$$

where the density bias, $b(\rho)$, entering equation (5) will later be shown to obey equation (42).

Similarly, the correlations between the counts in different bins of density can be investigated. The numbers N_1 of spheres with density in $\Delta_1 = [\rho_1 - \Delta\rho/2, \rho_1 + \Delta\rho/2]$ and N_2 of spheres with density in $\Delta_2 = [\rho_2 - \Delta\rho/2, \rho_2 + \Delta\rho/2]$ are unbiased and have a covariance

$$\langle N_1 N_2 \rangle = \bar{N}_1 \bar{N}_2 (1 + \xi b_1 b_2), \quad (12)$$

with b_1 and b_2 being defined as in equation (11) for the bins $\rho_1 \pm \Delta\rho/2$ and $\rho_2 \pm \Delta\rho/2$. The proof of this result is again derived in Appendix C and can be easily generalized to any number of concentric cells.

The consequence for the error budget of the one-cell PDF is as follows. Let us define $\hat{\mathcal{P}}(\rho_i) = N_i/N_t/\Delta\rho$, the estimate of the PDF measured from a set of N_t spheres when the range of densities is divided in bins centred on ρ_i with width $\Delta\rho$. First, this estimator, $\hat{\mathcal{P}}$, is unbiased. Equations (10) and (12) also yield the expected error on the estimate of the PDF

$$\langle \hat{\mathcal{P}}(\rho_i)^2 \rangle - \langle \hat{\mathcal{P}}(\rho_i) \rangle^2 = \frac{\bar{\mathcal{P}}(\rho_i)}{\Delta\rho N_t} + b_i^2 \xi (\bar{\mathcal{P}}(\rho_i))^2, \quad (13)$$

where the mean PDF in the bin is $\bar{\mathcal{P}}(\rho_i) = \int_{\Delta_i} \mathcal{P}(\rho) d\rho / \Delta\rho$ and $b_i^2 (\bar{\mathcal{P}}(\rho_i))^2$ is the mean value squared of the bias in the density bin, $(\int_{\Delta_i} \mathcal{P}(\rho) b(\rho) d\rho / \Delta\rho)^2$. Furthermore, the typical correlation between two distinct bins, $i \neq j$, is given by

$$\langle \hat{\mathcal{P}}(\rho_i) \hat{\mathcal{P}}(\rho_j) \rangle = \bar{\mathcal{P}}(\rho_i) \bar{\mathcal{P}}(\rho_j) (1 + \xi b_i b_j). \quad (14)$$

In particular, it is straightforward to see that equation (14) is fully consistent with equation (9) in the one-cell case. Indeed, from equations (14) and (13), one can compute for

instance the correlation between the estimated moment of order p of the density and the moment of order q

$$\begin{aligned} M_{pq} &= \langle \overline{\rho^p \rho^q} \rangle, \\ &= \sum_{i,j} (\Delta\rho)^2 \langle \hat{\mathcal{P}}(\rho_i) \hat{\mathcal{P}}(\rho_j) \rangle \rho_i^p \rho_j^q, \\ &= \sum_{i,j} (\Delta\rho)^2 \bar{\mathcal{P}}(\rho_i) \bar{\mathcal{P}}(\rho_j) (1 + \xi b_i b_j) \rho_i^p \rho_j^q + \sum_i \frac{\Delta\rho}{N_t} \rho_i^{p+q} \bar{\mathcal{P}}(\rho_i), \\ &= \frac{1}{N_t} \langle \rho^{p+q} \rangle + \langle \rho^p \rangle \langle \rho^q \rangle + \xi \langle b(\rho) \rho^p \rangle \langle b(\rho) \rho^q \rangle, \end{aligned}$$

which in terms of cumulants can be rewritten as

$$C_{pq} = \langle \overline{\rho^p \rho^q} \rangle_c = \frac{1}{N_t} \langle \rho^{p+q} \rangle_c + \xi \langle b(\rho) \rho^p \rangle_c \langle b(\rho) \rho^q \rangle_c, \quad (15)$$

so that equation (9) is recovered in the one-cell case.

The rest of the paper is devoted to demonstrating and validating equation (5).

3 TREE ORDER GENERATING FUNCTION

Let us first compute the generating function, φ_b of the cumulants containing any power of the densities in one location and one power of the density at an arbitrary distance. As we shall see, such cumulants enter the derivation of equation (5). For that purpose, we first consider n concentric cells in one location of space.

3.1 Definitions and relation to spherical collapse

Bernardeau et al. (2014) (hereafter BPC) computed $\mathcal{P}(\{\rho_k\})$, the joint one-point PDF of the density within concentric spheres, in a highly symmetric configuration (spherical symmetry) where non-linear solutions to the gravitational dynamical equations are known explicitly. The corresponding symmetry implies that the most likely dynamics (amongst all possible mappings between the initial and final density field) is that corresponding to spherical collapse. In the limit of small variance, BPC showed using a saddle approximation that the Laplace transform of $\mathcal{P}(\{\rho_k\})$ corresponds to the cumulant generating function of densities in concentric cells $\varphi(\{\lambda_k\})$, and can be predicted analytically. This function is indeed closely related to the non-linear evolution of a spherically symmetric perturbation in the linear growing mode regime and reads

$$\varphi(\{\lambda_k\}) = \sum_{p_i=0}^{\infty} \langle \Pi_i \rho_i^{p_i}(R_i) \rangle_c \frac{\Pi_i \lambda_i^{p_i}}{\Pi_i p_i!}, \quad (16)$$

where ρ_i is the density (in units of the average density) within the radius R_i . For this construction, it is essential that the cells are all spherical and concentric.

Let us denote $\zeta(\tau)$ the non-linear transform of the density so that

$$\rho = \zeta(\tau), \quad (17)$$

where ρ is the density within the radius R and τ is the linear density contrast within the radius $R\rho^{1/3}$ (for mass conservation). An explicit possible fit for $\zeta(\tau)$ is given by

$$\zeta(\tau) = \frac{1}{(1 - \tau/\nu)^\nu}, \quad (18)$$

where ν can be adjusted to the actual values of the cosmological parameters ($\nu = 21/13$ provides a good description of the spherical dynamics for an Einstein-de Sitter background for the range of τ values of interest). The main result of BPC was that the cumulant generating function at tree order could be computed explicitly from its Legendre transform², $\Psi(\{\rho_k\})$, as

$$\varphi(\{\lambda_k\}) = \sum_i \lambda_i \rho_i - \Psi(\{\rho_k\}), \quad (19)$$

where the $\{\rho_k\}$ are functions of the $\{\lambda_i\}$ via the stationary conditions

$$\lambda_i = \frac{\partial}{\partial \rho_i} \Psi(\{\rho_k\}), \quad i = 1 \cdots n. \quad (20)$$

Note that this condition can only be inverted as long as $\det \partial_i \partial_j \Psi \neq 0$ which defines a critical n -dimensional surface³. Under the above mentioned assumptions the rate function $\Psi(\{\rho_k\})$ entering equation (19) is explicitly written in term of the initial conditions as

$$\Psi(\{\rho_k\}) = \frac{1}{2} \sum_{ij} \Xi_{ij}(\{\rho_k\}) \tau_i(\rho_i) \tau_j(\rho_j), \quad (21)$$

where $\Xi_{ij}(\{R_k \rho_k^{1/3}\})$ is the inverse matrix of $\Sigma_{ij} = \sigma^2(R_i \rho_i^{1/3}, R_j \rho_j^{1/3})$, the initial cross correlation matrix of the densities computed at radii $R_i \rho_i^{1/3}$ and $R_j \rho_j^{1/3}$. The covariance matrix σ_{ij} encodes all dependency with respect to the initial power spectrum.

Note that equation (19) can be inverted as

$$\Psi(\{\rho_k\}) = \sum_i \lambda_i \rho_i - \varphi(\{\lambda_k\}), \quad (22)$$

with

$$\rho_i = \frac{\partial}{\partial \lambda_i} \varphi(\{\lambda_k\}), \quad i = 1 \cdots n. \quad (23)$$

The stationary condition then gives the expression of ρ_i as a function of the variables $\{\lambda_k\}$. Such a solution can be equivalently expressed in terms of the corresponding values of τ_i .

3.2 The $n+1$ cell formalism

Let us now consider the formal derivation of φ_b , the “one external leg” generating function of joint cumulants for $n+1$ cells centred on the same point when the $n+1$ th radius, $R_{n+1} = r_e$, is set apart (at this stage there is no assumption on the relative size of these radii). This configuration is

² Recently, [Bernardeau & Reimberg \(2015\)](#) showed that the prediction for the cumulant generating function given by equation (19) originates from a regime of large deviations (see [Touchette 2011](#), for a review) at play in the gravitational evolution of cosmic structures.

³ The critical surface defined by $\det \partial_i \partial_j \Psi = 0$ is at finite distance from the origin ensuring that the cumulant generating function has a non-zero radius of convergence. As a consequence, and contrary to what happens in the case of single or multivariate log-normal distributions ([Coles & Jones 1991](#); [Carron 2011](#); [Carron & Neyrinck 2012](#); [Carron & Szapudi 2015](#)), the density PDF here decays exponentially and Carleman’s criterion ensures that it is uniquely defined from its moments (see for instance [Akhiezer 1965](#)).

illustrated in the left-hand panel of Fig. 2 and is of particular interest since we will show in this section how it can be used to predict some configurations of the two-point statistics without any assumption on the separation. Later, we will also use it as a building block of the large-separation approximation of the two-point correlation function of concentric densities (see Section 4.2). This generating function simply reads

$$\varphi_b(\{\lambda_k\}; < r_e) = \sum_{p_i=0}^{\infty} \langle \rho(r_e) \prod_{i=1}^n \rho_i^{p_i}(R_i) \rangle_c \frac{\prod_{i=1}^n \lambda_i^{p_i}}{\prod_{i=1}^n p_i!}, \quad (24)$$

where $\rho(r_e)$ enters the cumulant $\langle \rho(r_e) \prod_{i=1}^n \rho_i^{p_i}(R_i) \rangle_c$ only as a linear power. Equation (24) is the generating function of the cumulants containing one power of the outer density and arbitrary powers of the n inner densities. It simply corresponds to the first derivative of the cumulant generating functions for $n+1$ cells taken at the origin

$$\varphi_b(\{\lambda_k\}; < r_e) = \frac{\partial}{\partial \lambda_{n+1}} \varphi(\lambda_1, \dots, \lambda_{n+1}) \Big|_{\lambda_{n+1}=0}. \quad (25)$$

Taking advantage of the stationary condition (23) applied to λ_{n+1} , we also have

$$\varphi_b(\{\lambda_k\}; < r_e) = \rho_{n+1}(\lambda_1, \dots, \lambda_n, 0), \quad (26)$$

where ρ_{n+1} is in turn computed in terms of the λ_i from the set of stationary conditions (20). Finally, equation (26) can also be re-expressed via equation (17) in terms of the corresponding linear density contrast as

$$\varphi_b(\{\lambda_k\}; < r_e) = \zeta(\tau(r_e)), \quad (27)$$

where $\tau(r_e) \equiv \tau_{n+1}(\lambda_1, \dots, \lambda_n, 0)$ is to be computed as a function of $\{\lambda_k\}$ for the specific case where λ_{n+1} is set to 0. We can then take advantage of decimation (see Appendix A, equation (A14)) to write $\tau(r_e)$ via the implicit equation

$$\tau(r_e) = \sum_{i=1}^n \sigma^2(r_e \zeta(\tau(r_e))^{1/3}, R_i \zeta(\tau_i)^{1/3}) \times \sum_{j=1}^n \Xi_{ij}(\{R_k \zeta(\tau_k)^{1/3}\}) \tau_j, \quad (28)$$

where the tensor and vector quantities (Ξ_{ij} , τ_i) are computed when only the first n cells are considered (so that the set $\sum_{j=1, n} \Sigma_{ij} \Xi_{jk} = \delta_{ik}$ together with the stationary conditions form a set of n coupled equations only). Technically, equation (28) can be solved given the values of $\{\tau_k\}_{k=1, \dots, n}$ which in turn can be expressed in terms of the variables $\{\lambda_k\}_{k=1, \dots, n}$.

Now note that equations (27) and (28) can be used to get the cumulant generating functions for any quantities linearly related to the density. In particular, the density in an infinitesimal shell at a given distance r_e reads $\rho(r_e < r < r_e + dr_e) = d\rho(r_e) r_e^3 / dr_e^3$ so that the corresponding cumulant generating function, $\varphi_b(\{\lambda_k\}; r_e)$, can be written as

$$\varphi_b(\{\lambda_k\}; r_e) = \frac{1}{r_e^3} \frac{d}{dr_e} \left(\frac{r_e^3}{3} \varphi_b(\{\lambda_k\}; < r_e) \right). \quad (29)$$

Thanks to rotational invariance, the value of the cumulant within an infinitesimal shell at a distance r_e is the same as if the density was computed at a distance r_e in any direction.

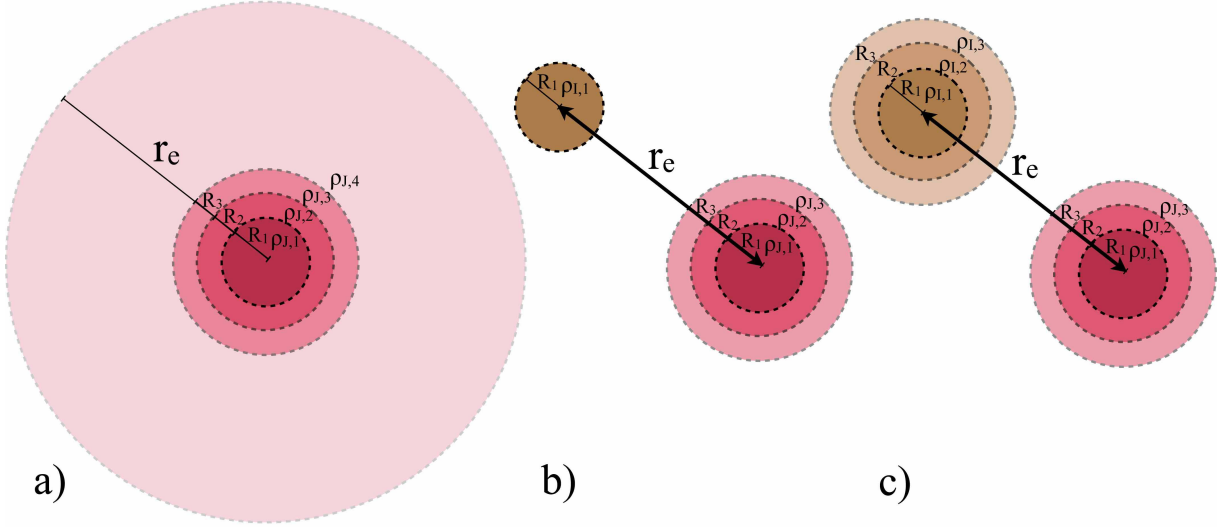


Figure 2. a) The configuration of 3 + 1 spherical cells in one location (red cells shown in the left-hand panel) can be used to compute the joint cumulants involving any power of the density in 3 concentric cells in one location (red cells displayed in the b) panel) and one power of the density in one cell (coloured in brown in the middle panel) at some arbitrary distance r_e from the rest, as described in Section 3.2. Those cumulants are the building blocks of the two-point PDF of concentric densities in the large-separation limit (see equation (38)). The corresponding configuration with $n = 3$ concentric cells in one location (red) and $m = 3$ concentric cells at a distance r_e (brown) is displayed in the c) panel.

Therefore, equation (29) also describes the cumulant generating function of concentric densities in spheres of radii R_i ($1 \leq i \leq n$) and density at some given distance r_e .

Finally note that the domain for ρ_e does not need to be a spherical cell and equation (29) can subsequently be integrated in any domain \mathcal{S} of arbitrary shape

$$\varphi_b(\{\lambda_k\}; \mathcal{S}) = \frac{1}{V_S} \int_{\mathcal{S}} d^3\mathbf{r} \varphi_b(\{\lambda_k\}; r), \quad (30)$$

with V_S the volume of the domain. This configuration is illustrated in the middle panel of Fig. 2.

4 BIAS IN THE LARGE-SEPARATION LIMIT

In the large-separation limit, $r_e \gg R_{\max} = \max_j R_j$, the results of the previous section can be pursued further to investigate the effects of cosmic variance on the measurement of the statistical properties of the field at scales R_j within a much larger survey.

4.1 Bias of the $n + 1$ cumulant generating function

In the large-separation limit where the internal radii, R_i , are all smaller than r_e and for realistic power spectra on cosmological scales, the cross-correlations $\sigma(r_e \zeta(\tau(r_e))^{1/3}, R_i \zeta(\tau_i)^{1/3})$ are much smaller than any internal moments $\sigma(R_i \zeta(\tau_i)^{1/3}, R_j \zeta(\tau_j)^{1/3})$. This property⁴ implies that in equation (28) the coefficients $\sigma^2(r_e \zeta(\tau(r_e))^{1/3}, R_i \zeta(\tau_i)^{1/3}) \Xi_{ij}$ are smaller than unity and therefore $\tau(r_e)$ is also small. The leading order expression of

φ_b can then be obtained using a Taylor expansion around $\tau(r_e) \approx 0$ of equation (18) into equations (27) and (28) so that

$$\varphi_b(\{\lambda_k\}; < r_e) = 1 + \sum_{i=1}^n \sigma^2(r_e, R_i \zeta(\tau_i)^{1/3}) \sum_{j=1}^n \Xi_{ij} \tau_j. \quad (31)$$

Pursuing this approximation further, we also expect that $\sigma(r_e, R_i \zeta(\tau_i)^{1/3})$ are essentially all equal and given by

$$\sigma(r_e, R_i \zeta(\tau_i)^{1/3}) \approx \sigma(r_e, 0) \equiv \sigma(< r_e). \quad (32)$$

Equation (31) can then be simplified as

$$\varphi_b(\{\lambda_k\}; < r_e) = 1 + \sigma^2(< r_e) \sum_{i=1}^n \sum_{j=1}^n \Xi_{ij} \tau_j. \quad (33)$$

In particular, this implies, via equation (29), that the cumulant generating function for the density at some large distance r_e from the n cells obeys

$$\varphi_b(\{\lambda_k\}; r_e) = 1 + \xi(r_e) \sum_{i=1}^n \sum_{j=1}^n \Xi_{ij} \tau_j, \quad (34)$$

where $\xi(r_e)$ is the dark matter correlation function at distance r_e

$$\xi(r_e) \equiv \frac{1}{r_e^2} \frac{d}{dr_e} \left(\frac{r_e^3}{3} \sigma^2(< r_e) \right). \quad (35)$$

Let us then define the bias cumulant generating function

$$b_\varphi(\{\lambda_k\}) \equiv \sum_{i=1}^n \sum_{j=1}^n \Xi_{ij} \tau_j, \quad (36)$$

so that

$$\varphi_b(\{\lambda_k\}; r_e) = 1 + \xi(r_e) b_\varphi(\{\lambda_k\}). \quad (37)$$

Using equation (36), the bias cumulant generating function can in principle be computed by means of equations (17),

⁴ This hierarchy of moments can be investigated by varying the spectral index and radii in equation (65) and shown to be valid for the range of power spectra and radii of cosmological interest.

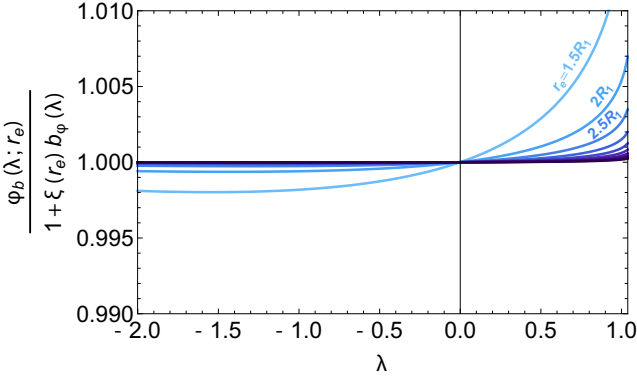


Figure 3. Ratio between $\varphi_b(\lambda; r_e)$, the exact generating function of cumulants of the form $\langle \rho_1^p \rho(r_e) \rangle_c$ and its large-distance approximation $\varphi_b(\lambda; r_e) = 1 + \xi(r_e) b_\varphi(\lambda)$ for a power-law density power spectrum with index $n_s = -1.6$ and variance $\sigma^2 = 0.3$. The separation r_e spans the range between $r_e = 3/2 R_1$ (light blue) and $r_e = 6 R_1$ (dark blue) as labelled. Note that the critical point in this case is $\lambda_c \approx 1.05$.

(20) and the inverse of the cross correlation matrix of the densities defined near (21) for any number of cells.

Within this approximation we see, recalling equation (24), that all cumulants of the form $\langle \rho(r_e) \rho_1^{p_1} \dots \rho_n^{p_n} \rangle_c$, are proportional to $\xi(r_e)$, where the $\rho_i(R_i)$ and $\rho(r_e)$ are located on cells centred at distance $r_e \gg R_i$ from one another.

For $n = 1$ cell, Fig. 3 shows the convergence of the approximation defined by equation (37) towards the exact value of $\varphi_b(\lambda; r_e)$ computed from equations (18), (27) and (28) as the distance r_e increases. The convergence is fast, reaching sub-percent precision even for relatively small separations $r_e \gtrsim 2 R_1$. As expected, the agreement is best for small values of λ , while deviations appear in the tails. Those conclusions are expected to hold similarly for a higher number n of cells.

4.2 The $n+m$ formalism at large separation

Let us finally focus on the $n + m$ formalism where n concentric cells at a distance r_e from m other concentric cells are considered. The rest of the paper is devoted to the study of this configuration, illustrated in the right-hand panel of Fig. 2, in order to investigate the large-separation two-point statistics of densities in concentric cells.

In this large-separation limit, we can write down (see appendix B) the joint cumulants for two sets, separated by r_e much larger than R_i , at leading order in terms of the two-point cumulants containing one power of the density in one location and any powers of the density in the second location (and generated by φ_b , see Section 3.2)

$$\langle \rho_1^{p_1} \dots \rho_n^{p_n} \rho_1^{q_1} \dots \rho_m^{q_m} \rangle_c = \frac{1}{\xi(r_e)} \langle \rho_1^{p_1} \dots \rho_n^{p_n} \rho'_i \rangle_c \langle \rho_i \rho_1^{q_1} \dots \rho_m^{q_m} \rangle_c, \quad (38)$$

where the $\{\rho_k = \rho_{1,k}\}_k$ correspond to a set of radii $\{R_k\}_k$ of cells centred at, say, the origin, the $\{\rho'_k = \rho_{2,k}\}_k$ correspond to a set of radii $\{R'_k\}_k$ of cells centred on a point at distance r_e from the origin and p_i and q_i are non-zero. Equation (38) can be easily understood from a diagrammatic point of view.

Indeed, in the large-separation regime, the dominant contribution will come from the configuration in which there is only one leg linking two connected diagrams belonging to each location (see also Bernardeau (1996) for details).

At the level of the generating functions, equations (37) and (38) imply that

$$\varphi(\{\lambda_k\}, \{\lambda'_k\}; r_e) = \varphi(\{\lambda_k\}) + \varphi(\{\lambda'_k\}) + \xi(r_e) b_\varphi(\{\lambda_k\}) b_\varphi(\{\lambda'_k\}), \quad (39)$$

where $\varphi(\{\lambda_k\}, \{\lambda'_k\}; r_e)$ is the generating functions of the joint cumulants $\langle \rho_1^{p_1} \dots \rho_n^{p_n} \rho_1^{q_1} \dots \rho_m^{q_m} \rangle_c$. Equation (39) is the cornerstone of this paper.

4.3 Consequences for the joint PDFs

The structure of equation (39) for the cumulant generating function has direct consequences at the level of the corresponding joint PDF. Let us consider again two sets of concentric cells separated by a distance r_e , and define the corresponding densities $\{\hat{\rho}_k\} \equiv \{\hat{\rho}_{1,k}\}$ and $\{\hat{\rho}'_k\} \equiv \{\hat{\rho}_{2,k}\}$. For large separations, the joint PDF $\mathcal{P}(\{\hat{\rho}_k\}, \{\hat{\rho}'_k\}; r_e)$ takes, at leading order, the following form

$$\mathcal{P}(\{\hat{\rho}_k\}, \{\hat{\rho}'_k\}; r_e) = \mathcal{P}(\{\hat{\rho}_k\}) \mathcal{P}(\{\hat{\rho}'_k\}) [1 + \xi(r_e) b(\{\hat{\rho}_k\}) b(\{\hat{\rho}'_k\})], \quad (40)$$

given that

$$\mathcal{P}(\{\hat{\rho}_k\}) = \int \frac{d\lambda_1}{2\pi i} \dots \frac{d\lambda_n}{2\pi i} \exp(-\lambda_i \hat{\rho}_i + \varphi(\{\lambda_k\})), \quad (41)$$

and

$$b(\{\hat{\rho}_k\}) \mathcal{P}(\{\hat{\rho}_k\}) = \int \frac{d\lambda_1}{2\pi i} \dots \frac{d\lambda_n}{2\pi i} b_\varphi(\{\lambda_k\}) \exp(-\lambda_i \hat{\rho}_i + \varphi(\{\lambda_k\})), \quad (42)$$

where the bias cumulant generating function $b_\varphi(\{\lambda_k\})$ is given by equation (36) and we have introduced the corresponding effective bias function, $b(\{\hat{\rho}_k\})$ (that will also simply be called “bias” or “bias function” in what follows). Note that measured densities will henceforth be denoted with a hat in order to avoid confusion with variables intervening in the computation of the cumulant generating functions (see for instance equation (19)).

Equation (42) is one of the main results of this paper. It defines the bias functions we introduced in equation (5), which is dual to equation (36) in the duality defined by PDF versus generating functions. Correspondingly, equation (40) is dual to equation (39). It is easy to show that the bias functions, $b(\{\hat{\rho}_k\})$, have an analytical asymptote at low density which can be derived using a steepest descent method in equations (41) and (42)

$$b(\{\hat{\rho}_k\}) \approx b_\varphi(\{\lambda_k = \partial_k \Psi(\{\hat{\rho}_i\})\}), \quad (43)$$

where $\partial_k \Psi = \partial \Psi / \partial \rho_k$. In practice, we will make use below of the following low-density approximation

$$b(\{\hat{\rho}_k\}) \approx \sum_{i=1}^n \sum_{j=1}^n \Xi_{ij}(\{\hat{R}_k \hat{\rho}_k^{1/3}\}) \zeta^{-1}(\hat{\rho}_j), \quad (44)$$

which allows to go beyond the critical point where $\lambda_i = \partial_i \Psi(\{\hat{\rho}_k\})$ is ill-defined.

4.4 Properties of the density bias

From equation (36), one can show that $b_\varphi = 0$ and $b'_\varphi(0) = 1$. Using equation (42), it follows that the bias functions obey the following two relations

$$\int_0^\infty d\hat{\rho}_k \mathcal{P}(\{\hat{\rho}_k\}) b(\{\hat{\rho}_k\}) = 0, \quad (45)$$

$$\int_0^\infty d\hat{\rho}_k \mathcal{P}(\{\hat{\rho}_k\}) b(\{\hat{\rho}_k\}) \hat{\rho}_k = 1. \quad (46)$$

These properties ensure the normalisation of the PDF and the definition of ξ , the dark matter correlation function.

In particular, we will make use of these identities to measure the density bias function either using the auto-correlation of cells of a given density

$$1 + b^2(\hat{\rho})\xi(r_e) = \frac{\int_0^\infty d\hat{\rho}_k d\hat{\rho}'_k \mathcal{P}(\{\hat{\rho}_k\}, \{\hat{\rho}'_k\}; r_e) \delta_D(\hat{\rho}_k - \hat{\rho}) \delta_D(\hat{\rho}'_k - \hat{\rho})}{\int_0^\infty d\hat{\rho}_k d\hat{\rho}'_k \mathcal{P}(\{\hat{\rho}_k\}) \mathcal{P}(\{\hat{\rho}'_k\}) \delta_D(\hat{\rho}_k - \hat{\rho}) \delta_D(\hat{\rho}'_k - \hat{\rho})}, \quad (47)$$

or their cross-correlations

$$1 + b(\hat{\rho})\xi(r_e) = \frac{\int_0^\infty d\hat{\rho}_k d\hat{\rho}'_k \mathcal{P}(\{\hat{\rho}_k\}, \{\hat{\rho}'_k\}; r_e) \delta_D(\hat{\rho}_k - \hat{\rho}) \delta_D(\hat{\rho}'_k - \hat{\rho})}{\int_0^\infty d\hat{\rho}_k \mathcal{P}(\{\hat{\rho}_k\}) \delta_D(\hat{\rho}_k - \hat{\rho})}. \quad (48)$$

In practice, measurements will be done in bins of a given width, meaning that Dirac delta functions will be replaced by stepwise functions (see Section 6.1 below).

4.5 Slope bias

In the two-cell case, instead of $(\hat{\rho}_1, \hat{\rho}_2)$, we will use the variables $(\hat{\rho}, \hat{s})$ describing the inner density $\hat{\rho} = \hat{\rho}_1$ and slope $\hat{s} = (\hat{\rho}_2 - \hat{\rho}_1)R_1/\Delta R$ with $\Delta R = R_2 - R_1$. The slope bias, $b(\hat{s})$, will be investigated using the relation

$$1 + b(\hat{s})\xi(r_e) = \frac{\int_0^\infty d\hat{\rho} \int_{-\infty}^\infty d\hat{s}' \mathcal{P}(\{\hat{\rho}\}, \{\hat{s}'\}; r_e) \delta_D(\hat{s}' - \hat{s}) \hat{\rho}}{\int_{-\infty}^\infty d\hat{s}' \mathcal{P}(\hat{s}') \delta_D(\hat{s}' - \hat{s})}. \quad (49)$$

In equation (49), $\mathcal{P}(\{\hat{\rho}\}, \{\hat{s}'\}; r_e)$ is a marginal of the two-cell PDF given by $\mathcal{P}(\{\hat{\rho}_1, \hat{s}_1\}, \{\hat{\rho}_2, \hat{s}_2\}; r_e) \delta_D(\hat{\rho}_1 - \hat{\rho}) \delta_D(\hat{s}_2 - \hat{s}')$ integrated over $\hat{\rho}_1, \hat{s}_1, \hat{\rho}_2$ and \hat{s}_2 . The slope bias, $b(\hat{s})$, is defined as the Inverse Laplace transform of the slope bias cumulant generating function

$$b_s(\mu) \equiv b_\varphi \left(-\frac{R_1}{\Delta R} \mu, \frac{R_1}{\Delta R} \mu \right), \quad (50)$$

via

$$b(\hat{s}) \mathcal{P}(\hat{s}) = \int \frac{d\mu}{2\pi i} b_s(\mu) \exp(-\mu \hat{s} + \varphi_s(\mu)), \quad (51)$$

with the slope PDF defined as

$$\mathcal{P}(\hat{s}) = \int \frac{d\mu}{2\pi i} \exp(-\mu \hat{s} + \varphi_s(\mu)). \quad (52)$$

Equations (51) and (50) follow from the rewriting

$$\lambda_1 \rho_1 + \lambda_2 \rho_2 = (\lambda_1 + \lambda_2) \rho_1 + \frac{\Delta R}{R_1} \lambda_2 s, \quad (53)$$

$$= \lambda \rho + \mu s, \quad (54)$$

which makes explicit the relation between (λ_1, λ_2) and (λ, μ) and consequently the relation between the slope cumulant generating function, φ_s , and φ

$$\varphi_s(\mu) = \varphi \left(\lambda_1 = -\frac{R_1}{\Delta R} \mu, \lambda_2 = \frac{R_1}{\Delta R} \mu \right). \quad (55)$$

Imposing $\lambda = 0$ therefore allows us to marginalise over ρ and get the slope bias.

It has to be noted that the slope bias obeys similar constraints as the density bias namely

$$\int_{-\infty}^\infty d\hat{s} \mathcal{P}(\hat{s}) b(\hat{s}) = 0, \quad (56)$$

$$\int_{-\infty}^\infty d\hat{s} \mathcal{P}(\hat{s}) b(\hat{s}) \hat{s} = 0. \quad (57)$$

5 IMPLEMENTATION FOR REALISTIC P_K

In order to explicitly compute the bias functions for a realistic (Λ CDM) power spectrum, $P^{\text{lin}}(k)$, and compare those predictions to simulations, we follow BPC and approximate the variance

$$\sigma^2(R, R) = \int \frac{d^3 \mathbf{k}}{(2\pi)^3} P^{\text{lin}}(k) W_{3D}^2(kR), \quad (58)$$

using an analytical formula describing the scale-dependencies

$$\sigma^2(R, R) = \frac{2\sigma^2(R_p, R_p)}{(R/R_p)^{n_1+3} + (R/R_p)^{n_2+3}}, \quad (59)$$

where n_1 and n_2 are chosen to reproduce the linear theory index $n(R) = -3 - d \log(\sigma^2(R, R))/d \log R$ and running, $\alpha(R) = d \log(n(R))/d \log R$ at the pivot scale R_p . Note that in fine the amplitude, $\sigma_0 = \sigma(R_p, R_p)$, can be simply measured and not necessarily predicted by linear theory as it can be easily scaled out using the relation

$$\varphi_{\sigma_0}(\{\lambda_k\}) = \frac{1}{\sigma_0^2} \varphi_1(\{\lambda_k/\sigma_0^2\}), \quad (60)$$

while the variable ρ_k are independent of σ_0 . In practice, we take here $(n_1 + n_2)/2 = -1.58$ and $n_1 - n_2 = 1.23$ at $R_p = 10 \text{ Mpc}/h$ and $\sigma^2(R_p, R_p)$ is measured at different redshifts in the simulation.

Using the parametrization (59) of $\sigma(R, R)$, we can then analytically predict the one-cell rate function Ψ , the cumulant generating function φ and the bias function b_φ and finally numerically compute the PDF $\mathcal{P}(\hat{\rho})$ and bias function $b(\hat{\rho})$. In this paper, the numerical integrations are done, in the one-cell case, using a path parallel to the imaginary axis $\lambda = \lambda_0 + \Delta \lambda i$ where $\lambda_0 = \Psi'(\min(\hat{\rho}, \rho_c))$, i goes from 1 to the number of points n_{steps} which is set to 1000 and the step, $\Delta \lambda$, depends on the value of the density $\hat{\rho}$ (typically larger for smaller densities).

The resulting prediction for the density bias is shown in the left-hand panel of Fig. 4 (solid lines). As expected from previous studies, this figure shows that the bias is zero only in regions where the density is around the mean

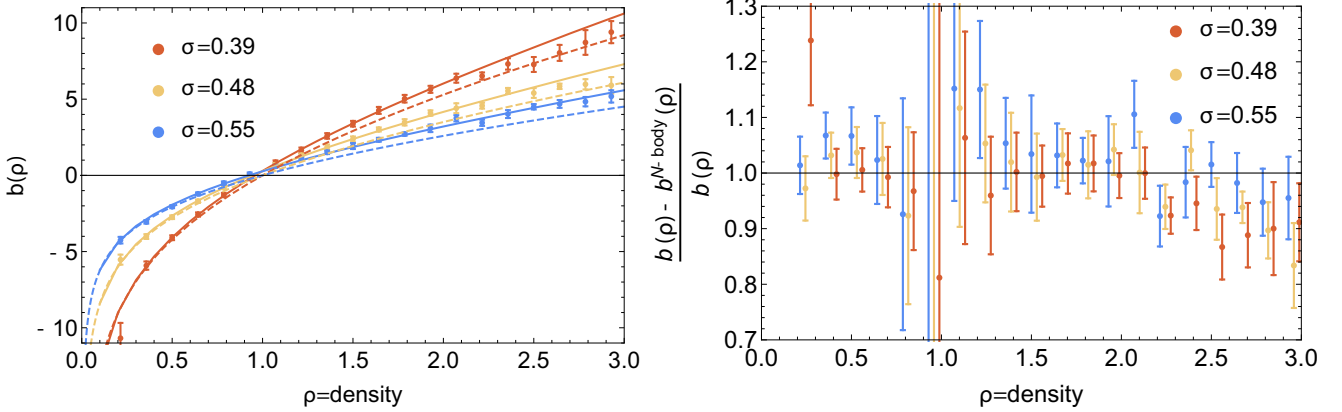


Figure 4. Left-hand panel: The density bias function $b(\hat{\rho})$ for different values of the variance ($\sigma=0.39, 0.48$ and 0.55 for respectively the red, yellow and blue lines). The prediction computed numerically via equation (42) is displayed with solid lines while the dashed lines correspond to the analytical low-density approximations given by equation (44). Measurements in the simulation for spheres of radius $R = 10\text{Mpc}/h$ separated by $r_e = 20\text{Mpc}/h$ are shown with error bars and are successfully compared to the numerical prediction (solid lines) for the full range of variances and densities. The low-density approximations break down for $\rho \gtrsim 1$. Right-hand panel: Residuals between the measured density bias (error bars) and its numerical prediction from equation (42) (displayed with solid lines on the left-hand panel). The yellow and red error bars have been shifted along the x-axis respectively by 0.03 and 0.06 for aesthetic purposes.

i.e. $\rho \approx 1$. This can easily be understood as $\mathcal{P}(\{\hat{\rho}\}, \{\hat{\rho}' = 1\}; r_e)/\mathcal{P}(1) = \mathcal{P}(\hat{\rho})$ if r_e is large enough. Overdense regions are then positively biased while underdense regions are less clustered (negatively biased). This is also consistent with Kaiser (1984) as the left-hand panel of Fig. 4 shows that the larger the density, the stronger the bias. Note that unlike the Gaussian peak bias which scales like the contrast for large densities, here the scaling of the non-linear bias departs from linearity with typically $b \propto \rho^{0.8}$ for large densities but is recovered in the very large density limit $\hat{\rho} \gtrsim 10$. Indeed, it is found (see Appendix E for more details) that the asymptotic behaviour of the bias function is given by

$$b(\hat{\rho}) \xrightarrow{\hat{\rho} \rightarrow \infty} \frac{(\hat{\rho} - \rho_c)(\lambda_c \rho_c (\rho_c^{1/\nu} - 1 - 1/\nu) + \frac{1}{\nu} \varphi_c)}{\nu(\rho_c^{1/\nu} - 1)^2 \rho_c^{1-1/\nu}} \quad (61)$$

when the integral is dominated by the singular point in equations (41) and (42). It has to be noted that this asymptotic behaviour depends on the existence of the singular point, ρ_c , which is due to the application of a large-deviation principle to the density field. However, as shown in Uhlemann et al. (2015), this singular point can be removed by applying a large-deviation principle to a non-linear transformation of the density field, for instance its logarithm. In that case, we expect the asymptotic behaviour to be slightly modified as shown in that paper.

Note also that the bias of large-density regions is reduced for larger variances. This trend persists even if the density bias is plotted against the contrast $\nu_c = (\hat{\rho} - 1)/\sigma$ instead of $\hat{\rho}$. The left-hand panel of Fig. 4 also displays the low-density approximation given by equation (44). This approximation is only valid in a limited range of densities typically for $\hat{\rho} \lesssim 1$.

In the two-cell case, one needs to compute the covariance matrix between initial densities in spheres of radii R_1 and R_2

$$\sigma^2(R_i, R_j) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} P^{\text{lin}}(k) W_{3D}(kR_i) W_{3D}(kR_j), \quad (62)$$

where W_{3D} is the top-hat filter function

$$W_{3D}(k) = \frac{3}{k^2} (\sin(k)/k - \cos(k)). \quad (63)$$

Again, for the sake of simplicity, we choose to parametrize this covariance matrix by

$$\sigma^2(R_i, R_i) = \sigma^2(R_p) \left(\frac{R_i}{R_p} \right)^{-n_s(R_p)-3}, \quad (64)$$

$$\sigma^2(R_i, R_{j>i}) = \sigma^2(R_p) R_p^{n_s(R_p)+3} \mathcal{G}(R_i, R_j, n_s(R_p)), \quad (65)$$

where

$$\begin{aligned} \mathcal{G}(x, y, n_s) &= R_p^{-n_s-3} \frac{\int d^3\mathbf{k} k^{n_s} W_{3D}(kx) W_{3D}(ky)}{\int d^3\mathbf{k} k^{n_s} W_{3D}(kR_p) W_{3D}(kR_p)} \\ &= \frac{(x+y)^\alpha (x^2+y^2-\alpha xy) - (y-x)^\alpha (x^2+y^2+\alpha xy)}{2^\alpha (n_s+1) x^3 y^3}, \end{aligned}$$

with $\alpha = 1 - n_s$. We are now in a position to compute the two-cell rate function, cumulant generating function, PDF and finally the two-cell bias function. The slope bias $b(\hat{s})$ predicted by this formalism is shown in Fig. 6 (right-hand panel). Two configurations are shown to be uncorrelated ($b(\hat{s}) = 0$) and roughly correspond to slopes $\hat{s} \approx \pm 0.5$ with some σ -dependence. This situation is of particular interest as, according to equation (13), cosmic variance is drastically reduced in this case where only subdominant contributions will appear such as Poisson noise and small-scale effects. Besides this noteworthy case, it is found that small slopes ($|\hat{s}| \lesssim 0.5$) are negatively biased while regions with larger (positive or negative) slope are more clustered. As expected, the bias is stronger for large slopes which typically correspond to sharp peaks (or voids) and will scale like $b(\hat{s}) \propto \hat{s}$ in the very large slope limit as explained in appendix F. This asymptotic large $|s|$ behaviour can be compared once again to the linear Kaiser bias, given by equation (1), but in the contrast of the peak rather than the slope. Note also that the asymmetry of the bias function is weakened with variance.

The predictions for the density and slope bias functions are compared against simulations in the next section.

6 VALIDATION AGAINST SIMULATIONS

In the following we present the measurement of this bias function for the density $\hat{\rho}$ and for the density slope \hat{s} .

The dark matter simulation (carried out with `Gadget2`, [Springel 2005](#)) is characterized by the following Λ CDM cosmology: $\Omega_m = 0.265$, $\Omega_\Lambda = 0.735$, $n = 0.958$, $H_0 = 70 \text{ km}\cdot\text{s}^{-1}\cdot\text{Mpc}^{-1}$ and $\sigma_8 = 0.8$, $\Omega_b = 0.045$ within one standard deviation of WMAP7 results ([Komatsu et al. 2011](#)). The box size is 500 Mpc/h sampled with 1024^3 particles, the softening length 24 kpc/h. Initial conditions are generated using `mpggrafic` ([Prunet et al. 2008](#)). An Octree is built to count efficiently all particles within a given sequence of concentric spheres of radii between $R = 4, 5 \dots$ up to 18 Mpc/h. The center of these spheres is sampled regularly on a grid of 10 Mpc/h aside, leading to $50^3 = 125,000$ estimates of the density per snapshot. Note that the cells overlap for radii larger than 5 Mpc/h.

6.1 Density bias function

The density bias is estimated from spheres of radius $R = 10 \text{ Mpc/h}$ that are separated by $r_e = 20 \text{ Mpc/h}$ using the cross-correlations defined in equation (48). More precisely, we compute a sum over each sphere I with density ρ_I and its 6 neighbours at distance $r_e = 20 \text{ Mpc/h}$ labelled with the indices $\alpha_{I,j}$ for $1 \leq j \leq 6$

$$\hat{b}(\hat{\rho}) = \frac{1}{\xi} \left[\frac{\sum_I \sum_{j=1}^6 \mathcal{B}(\hat{\rho} - \Delta\rho/2 \leq \rho_I \leq \hat{\rho} + \Delta\rho/2) \rho_{\alpha_{I,j}}}{6 \sum_I \mathcal{B}(\hat{\rho} - \Delta\rho/2 \leq \rho_I \leq \hat{\rho} + \Delta\rho/2)} - 1 \right]$$

where \mathcal{B} is a boolean function which evaluates to one if the density is in a bin centred on $\hat{\rho}$ with width $\Delta\rho = 3/21$ and the measured dark matter correlation function at distance r_e is given by

$$\hat{\xi}(r_e) = \frac{\sum_{I=1}^{N_t} \sum_{j=1}^6 \rho_I \rho_{\alpha_{I,j}}}{6 N_t} - 1. \quad (66)$$

In practice, we count all pairs of spheres only once when computing $\hat{\xi}(r_e)$ by only considering three neighbours for each sphere. Error bars are then evaluated by computing the error on the mean of the density bias when the simulation is divided into eight sub-cubes.

The density bias $\hat{b}(\hat{\rho})$ measured for different values of the variance is shown in the left-hand panel of Fig. 4 and successfully compared to the predictions obtained in Section 5. In particular and as expected, it is found that the bias is null for densities of 1. For count-in-cells, this trivially implies that imposing that the density equals its mean value within some sphere does not impact its neighbourhood sufficiently far enough. The right-hand panel of Fig. 4 displays the corresponding residuals and confirms the extremely good agreement between theory and measurements. Note that errors are relatively large for $\hat{\rho} \approx 1$ as $b(\hat{\rho})$ is very close to zero in this region. Note also that the apparent discrepancy between theory and prediction for the lower density bin at $\sigma = 0.39$ is mainly due to the fact that we did not take into account the size of the bin when predicting the density bias from equation (42) while it would have been necessary especially in such steep regions of the plot. Eventually, we conclude that the prediction is within the error bars of the simulations for the whole range of densities probed by our simulation meaning that a 10% accuracy at worst can be

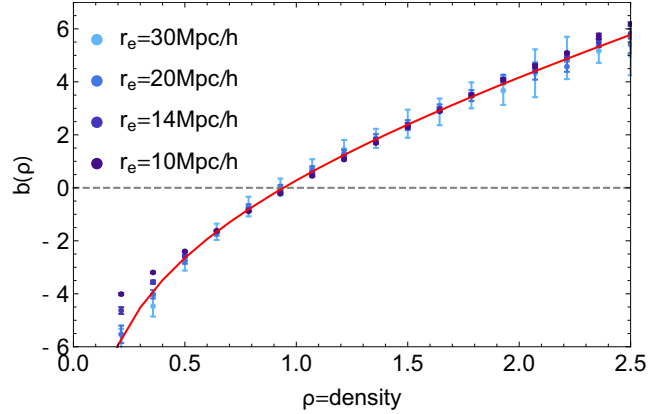


Figure 5. The density bias function $b(\rho)$ for $\sigma = 0.97$ measured in our simulation for different separations r_e as labelled. The large-separation prediction is displayed in red.

achieved with large-deviation theory. It is striking to see that the accuracy is poorer at higher redshift which could indicate that the main source of uncertainty here is not due to the theoretical prediction but to some numerical artefacts of our simulation. The (analytical) low-density approximation is shown to give a reasonable fit from $\approx 10\%$ precision for $\rho \lesssim 1$ to order one for larger densities.

In order to investigate the convergence towards the large-separation limit, we also measure the density bias at different separations r_e (see Fig. 5, right-hand panel). For separation larger than $2R_1 = 20 \text{ Mpc/h}$ (adjacent spheres!), the prediction lies within the error bars of the simulations for the full range of density probed ($0.2 < \hat{\rho} < 2.5$).

Note that appendix D shows that the density bias estimated from the auto-correlation (see equation 47) is fully consistent with the cross-correlation based estimator used in this section.

6.2 Slope bias function

In order to measure the slope bias in the simulation, we consider the set of concentric spheres with radii $R_1 = 10 \text{ Mpc/h}$ and $R_2 = 11 \text{ Mpc/h}$ and pairs that are separated by $r_e = 20 \text{ Mpc/h}$ so that each set I of two concentric spheres with density ρ_I and slope s_I has six neighbours at distance r_e . Following equation (49), we compute again a sum over each set I and its 6 neighbouring sets at distance $r_e = 20 \text{ Mpc/h}$ labelled with the indices $\alpha_{I,j}$ for $1 \leq j \leq 6$

$$\hat{b}(\hat{s}) = \frac{1}{\hat{\xi}} \left[\frac{\sum_{I,j} \mathcal{B}(\hat{s} - \Delta s/2 \leq s_I \leq \hat{s} + \Delta s/2) \rho_{\alpha_{I,j}}}{6 \sum_I \mathcal{B}(\hat{s} - \Delta s/2 \leq s_I \leq \hat{s} + \Delta s/2)} - 1 \right], \quad (67)$$

where the bin width is set to $\Delta s = 5/21$ here.

The slope bias function $b(\hat{s})$ measured for different values of the variance by means of equation (67) is shown in Fig. 6 and can be compared to the large-deviation prediction displayed with solid lines. The two uncorrelated configurations predicted by our model are recovered around $\hat{s} \approx \pm 0.5$. One could make use of this striking feature in order to minimize large-scale clustering while restricting counts to such slopes according to equation (12)⁵. Predictions and mea-

⁵ The correlation vanishes at leading order in our calculations,

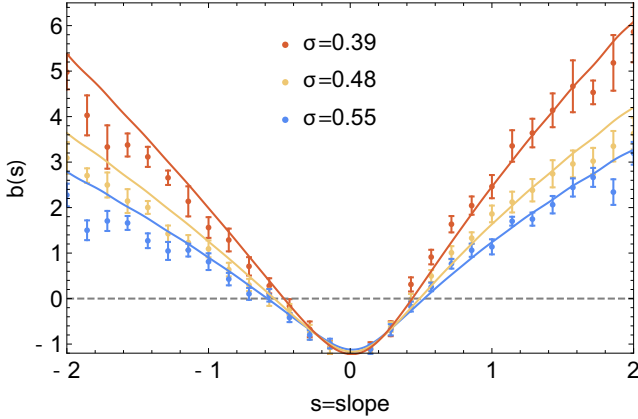


Figure 6. The slope bias $b(\hat{s})$ for different values of the variance as labelled. Measurements in the simulation for spheres of radii $R_1 = 10\text{Mpc}/h$ and $R_2 = 11\text{Mpc}/h$ separated by $r_e = 20\text{Mpc}/h$ using the estimator defined in equation (67) are displayed with error bars and compared with the prediction (solid line).

measurements are in good agreement for small slopes $|\hat{s}| \lesssim 1$ and both show that small slope regions are less clustered than sharper environments. They start to depart from one another in the tails of the distribution. Note that already the PDF of the slope shows some level of discrepancy in this regime, which can be due to our numerical integration, our choice of parametrization for the variance (which does not take into account the running of the spectral index) or some unforeseen artifacts from the simulation. Further work is necessary to understand exactly why those discrepancies arise. It is hoped that larger simulations and more robust analytical approximations (following Uhlemann et al. 2015) will resolve these issues.

Fig. 7 compares the measurements to the prediction for different values of the separation, $r_e = R_1$, $\sqrt{2}R_1$ (along a diagonal), $2R_1$ (adjacent cells) and $3R_1$. The convergence towards the large-separation prediction (in red) is quick since the prediction starts to deviate from the error bars of the measurements only for $r_e \lesssim 2R_1$ (adjacent cells!). Note that as expected, error bars are larger for larger separations.

7 APPLICATION TO COSMIC VARIANCE

We propose here to illustrate how the large-deviation prediction for the bias can be used to predict the statistics of errors. For that purpose, we measured the density PDF in 25^3 spheres of radius $R = 10\text{Mpc}/h$ at redshift $z = 1$ in our simulation. As already pointed out in Section 2.3, our estimate of the PDF can suffer from two types of errors : shot noise and cosmic variance. We use equation 13 to compute the expected error and display the result in Figure 8. For the particular configuration used here, the cosmic variance is dominant compared to the shot noise error. As expected, error bars obtained by resampling seem to slightly underestimate the error budget. If the distance between the spheres increases, the amount of cosmic variance decreases as the

not taken into account proximity effects and higher order correlation functions.

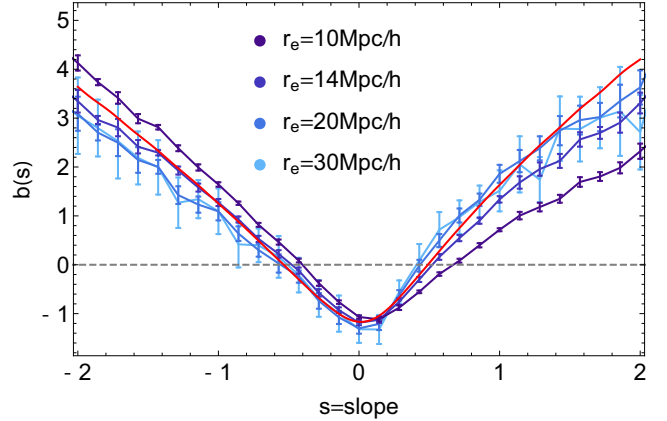


Figure 7. The slope bias $b(\hat{s})$ measured in the simulation at redshift $z = 0.97$ corresponding to $\sigma = 0.48$ and for different values of the separation from $r_e = R_1 = 10\text{Mpc}/h$ to $3R_1 = 30\text{Mpc}/h$ as labelled. The measurements quickly converge towards the predicted large-separation limit (red solid line).

spheres are less correlated. One can anticipate that for a given accessible volume, there is a balance to find between reducing the number of spheres to have them as independent as possible or increasing the number of spheres to reduce the shot noise. Finding the optimal number of spheres and radii to consider for a given survey geometry is left for future works.

8 CONCLUSION

We have shown how to compute joint statistics of the density within multiple concentric spheres in two regimes : i) the generating function of cumulants containing any powers of concentric densities in one location and one power of density at some *arbitrary* distance and ii) the two-point correlation function of the density in concentric spheres when the cells are sufficiently well apart. The latter allowed us to estimate the bias and cosmic variance involved in applying standard concentric count-in-cells statistics to cosmological fields of finite extent in that regime. The accuracy of the large-distance approximation was quantified against numerical simulations and shown to be valid even for adjacent cells ($\approx 20\text{Mpc}/h$ here), strengthening earlier findings by Bernardeau (1996) in the one-cell case. These simulations were also used to assess the validity of the large-deviation principle, which only formally holds in the zero variance limit, but was shown to give accurate predictions even for variance of order unity, in the so-called quasi-linear regime. The bias functions ($b(\hat{\rho})$, $b(\hat{s})$ in this work) allow us to quantify the covariances expected in finite volume effects, hence build accurate maximum likelihood estimators that could be applied to future surveys. In particular, the shape of $b(\hat{s})$ we found suggests that tailoring counts to cells which have slopes of the order of $\pm 1/2$ could be used to mitigate its effect. The formalism presented here can be straightforwardly extended to predict the bias of multiple concentric spheres including the bias of regions having a given density and slope $b(\hat{\rho}, \hat{s})$.

While the large-deviation predictions of count-in-cell bias functions rely on numerical integration in the complex

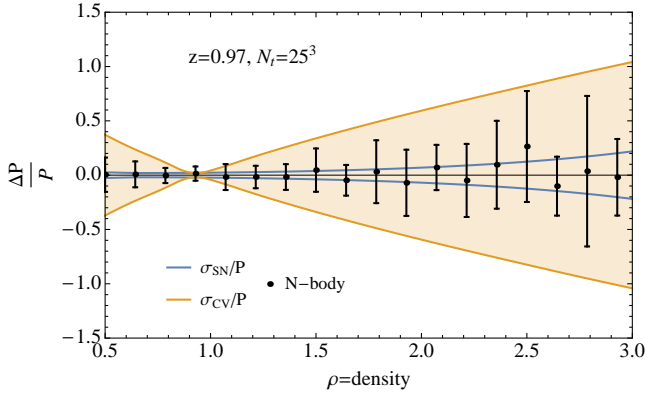


Figure 8. Expected shot noise (blue) and cosmic variance (yellow) for the density PDF measured from 25^3 spheres of radius $R = 10 \text{ Mpc}/h$ at redshift $z = 1$ regularly drawn in our simulation. Those predictions are compared with error bars (black) estimated as the standard deviation among eight sub-cubes of the simulation.

plane, the recent results of Uhlemann et al. (2015) sets the stage for accurate analytical approximations, using the logarithmic transform of the density. It would therefore be of interest to apply this log-transformation to the two-point statistics presented in this paper. This will be the topic of upcoming work.

It has to be emphasized that the scale-dependence of the count-in-cell bias (that is clearly seen in simulations on scales below $\approx 20 \text{ Mpc}/h$) can not be captured by the large-deviation principle used in this work. However, this formalism is a unique opportunity to get insights into the mildly non-linear evolution of halo biasing in contrast to peak models which usually assumes Gaussianity or very recently a Zel'dovich ballistic displacement of the initial peaks (Baldauf et al. (2016), Baldauf et al., in prep). An hybrid analysis between both approaches is challenging but would be of great interest. Building up this peak theory in the large-deviation regime is left for future works.

It would also be worth extending the investigation of biases and cosmic variance to the velocity field following the results obtained by Bernardeau (1992) in the one-point case, and to projected densities.

Acknowledgements: This work is partially supported by the grants ANR-12-BS05-0002 and ANR-13-BS05-0005 of the French *Agence Nationale de la Recherche*. The simulations were run on the *Horizon* cluster. We acknowledge support from S. Rouberol for running the cluster for us and warmly thank Cora Uhlemann for useful comments and careful reading of the manuscript. SC also thanks Marcelo Alvarez and Xin Wang for insightful discussions.

References

Abazajian K., Adelman-McCarthy J. K., Agüeros M. A., Allam S. S., Anderson S. F., Annis J., Bahcall N. A., Baldry I. K., Bastian S., Berlind A., et al. 2003, *AJ*, 126, 2081
 Akhiezer N., 1965, The classical moment problem: and some related questions in analysis. University mathematical monographs, Oliver & Boyd

Baldauf T., Codis S., Desjacques V., Pichon C., 2016, *MNRAS*, 456, 3985
 Baldauf T., Seljak U., Desjacques V., McDonald P., 2012, *Phys. Rev. D*, 86, 083540
 Balian R., Schaeffer R., 1989, *A&A*, 220, 1
 Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, *ApJ*, 304, 15
 Baugh C. M., Gaztanaga E., Efstathiou G., 1995, *MNRAS*, 274, 1049
 Bernardeau F., 1992, *ApJ Lett.*, 390, L61
 Bernardeau F., 1996, *A&A*, 312, 11
 Bernardeau F., Codis S., Pichon C., 2015, *MNRAS*, 449, L105
 Bernardeau F., Colombi S., Gaztanaga E., Scoccimarro R., 2002, *Phys. Rep.*, 367, 1
 Bernardeau F., Crocce M., Scoccimarro R., 2008, *Phys. Rev. D*, 78, 103521
 Bernardeau F., Pichon C., Codis S., 2014, *Phys. Rev. D*, 90, 103519
 Bernardeau F., Reimberg P., 2015, ArXiv e-prints
 Carron J., 2011, *ApJ*, 738, 86
 Carron J., Neyrinck M. C., 2012, *ApJ*, 750, 28
 Carron J., Szapudi I., 2015, ArXiv e-prints
 Catelan P., Porciani C., Kamionkowski M., 2000, *MNRAS*, 318, L39
 Coles P., Jones B., 1991, *MNRAS*, 248, 1
 Colombi S., Bouchet F. R., Schaeffer R., 1995, *ApJ Sup.*, 96, 401
 Crocce M., Castander F. J., Gaztanaga E., Fosalba P., Carretero J., 2015, *MNRAS*, 453, 1513
 Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371
 Dekel A., Lahav O., 1999, *ApJ*, 520, 24
 Dekel A., Rees M. J., 1987, *Nature*, 326, 455
 Fry J. N., 1996, *ApJ Lett.*, 461, L65
 Fry J. N., Gaztanaga E., 1993, *ApJ*, 413, 447
 Giannantonio T., Porciani C., 2010, *Phys. Rev. D*, 81, 063530
 Guo H., Jing Y. P., 2009, *ApJ*, 702, 425
 Kaiser N., 1984, *ApJ Lett.*, 284, L9
 Kaiser N., 1987, *MNRAS*, 227, 1
 Kiessling A., Cacciato M., Joachimi B., Kirk D., Kitching T. D., Leonard A., Mandelbaum R., Schäfer B. M., Sifón C., Brown M. L., Rassat A., 2015, *Space Sci. Rev.*, 193, 67
 Komatsu E., Smith K. M., Dunkley J., Bennett C. L., Gold B., Hinshaw G., Jarosik N., Larson D., et al. 2011, *ApJ Sup.*, 192, 18
 Laureijs R., Amiaux J., Arduini S., Augüeres J., Brinchmann J., Cole R., Cropper M., Dabin C., Duvet L., Ealet A., et al. 2011, ArXiv e-prints
 Lazanu A., Giannantonio T., Schmittfull M., Shellard E. P. S., 2015, ArXiv e-prints
 Levi M., Bebek C., Beers T., Blum R., Cahn R., Eisenstein D., Flaugher B., Honscheid K., Kron R., Lahav O., McDonald P., Roe N., Schlegel D., representing the DESI collaboration 2013, ArXiv e-prints
 Ludlow A. D., Porciani C., 2011, *MNRAS*, 413, 1961
 Lumsden S. L., Heavens A. F., Peacock J. A., 1989, *MNRAS*, 238, 293
 Manera M., Gaztanaga E., 2011, *MNRAS*, 415, 383
 Maurogordato S., Lachieze-Rey M., 1987, *ApJ*, 320, 13
 McDonald P., Roy A., 2009, *JCAP*, 8, 020
 Mirbabayi M., Schmidt F., Zaldarriaga M., 2015, *JCAP*, 7, 030
 Moster B. P., Somerville R. S., Newman J. A., Rix H.-W., 2011, *ApJ*, 731, 113
 Nusser A., Davis M., 1994, *ApJ Lett.*, 421, L1
 Pen U.-L., 1998, *ApJ*, 504, 601
 Prunet S., Pichon C., Aubert D., Pogossyan D., Teyssier R., Gottloeber S., 2008, *ApJ Sup.*, 178, 179
 Sánchez C., Carrasco Kind M., Lin H., Miquel R., Abdalla F. B., Amara A., 2014, *MNRAS*, 445, 1482

- Schneider A., Teyssier R., 2015, *JCAP*, 12, 049
 Scoccimarro R., Colombi S., Fry J. N., Frieman J. A., Hivon E., Melott A., 1998, *ApJ*, 496, 586
 Senatore L., 2015, *JCAP*, 11, 007
 Springel V., 2005, *MNRAS*, 364, 1105
 Szapudi I., Colombi S., 1996, *ApJ*, 470, 131
 Szapudi I., Colombi S., Bernardeau F., 1999, *MNRAS*, 310, 428
 Taruya A., Nishimichi T., Saito S., 2010, *Phys. Rev. D*, 82, 063522
 Tegmark M., Peebles P. J. E., 1998, *ApJ Lett.*, 500, L79
 Touchette H., 2011, ArXiv e-prints
 Uhlemann C., Codis S., Pichon C., Bernardeau F., Reimberg P., 2015, ArXiv e-prints
 Valageas P., 2002, *A&A*, 382, 412

APPENDIX A: RADII DECIMATIONS

The purpose of this appendix is to insure that the expression of the cumulant generating function $\varphi(\{\lambda\})$, given by equation (19), is consistent with variable decimation, i.e. we want to check that

$$\varphi(\{\lambda_1, \dots, \lambda_n\}) = \varphi(\{\lambda_1, \dots, \lambda_n, \lambda_{n+1} = 0, \dots, \lambda_{n+m} = 0\}), \quad (\text{A1})$$

where the left-hand side is computed from n cells whereas the right-hand side is computed with $n+m$ cells. This property was also studied in BPC but we propose here a slightly different proof and we will make use of some of the results – in particular equation (A14) – in the main text.

By mathematical induction, it is sufficient to prove equation (A1) when m is set to 1. We will therefore show here that the following property holds

$$\varphi(\{\lambda_1, \dots, \lambda_n\}) = \varphi(\{\lambda_1, \dots, \lambda_n, \lambda_{n+1} = 0\}), \quad (\text{A2})$$

where the left-hand side is still computed from n cells whereas the right-hand side is now computed with $n+1$ cells.

Let us define a set \mathcal{A} of n cells and a set \mathcal{B} of $m=1$ cell. One can then define the covariance matrix Σ defined by $\Sigma_{ij} = \sigma^2(R_i \zeta(\tau_i)^{1/3}, R_j \zeta(\tau_j)^{1/3})$ between two any cells of the union of \mathcal{A} and \mathcal{B} . The covariance matrix restricted to \mathcal{A} will be denoted $\tilde{\Sigma} = \Sigma|_{\mathcal{A}}$. For the sake of clarity, in this appendix, greek indices refer to the set \mathcal{A} and therefore go between 1 and n and a tilde is displayed over the corresponding operators while in $\mathcal{A} \cup \mathcal{B}$ we use roman indices which go between 1 and $n+1$ and no tilde.

The cumulant generating functions for the n cells in \mathcal{A} is given by

$$\varphi(\{\lambda_1, \dots, \lambda_n\}) = \lambda_\mu \tilde{\rho}_\mu - \frac{1}{2} \tilde{\Xi}_{\mu\nu} \tilde{\tau}_\mu \tilde{\tau}_\nu, \quad (\text{A3})$$

with the stationary conditions

$$\lambda_\kappa = \frac{\partial \tilde{\Psi}}{\partial \tilde{\rho}_\kappa} = \tilde{\Xi}_{\mu\kappa} \tilde{\tau}_\mu + \frac{1}{2} \frac{\partial \tilde{\Xi}_{\mu\nu}}{\partial \tilde{\rho}_\kappa} \tilde{\tau}_\mu \tilde{\tau}_\nu, \quad (\text{A4})$$

where $\tilde{\Xi}$ is the inverse matrix of $\tilde{\Sigma}$ and we use implicit summations over repeated indices μ, ν between 1 and n .

Our purpose is to show that it is identical to the expression of $\varphi(\{\lambda_i\}_{1 \leq i \leq n+1})$ describing the cumulant generating function of the $n+1$ cells when the last value λ_{n+1} is set to zero. In this latter case we have

$$\varphi(\{\lambda_1, \dots, \lambda_n, 0\}) = \lambda_\mu \rho_\mu - \frac{1}{2} \Xi_{ij} \tau_i \tau_j, \quad (\text{A5})$$

with the stationary conditions

$$\lambda_\kappa = \Xi_{i\kappa} \tau_i \frac{d\tau_\kappa}{d\rho_\kappa} + \frac{1}{2} \frac{\partial \Xi_{ij}}{\partial \rho_\kappa} \tau_i \tau_j; \quad (\text{A6})$$

$$\lambda_{n+1} = 0 = \Xi_{n+1i} \tau_i \frac{d\tau_{n+1}}{d\rho_{n+1}} + \frac{1}{2} \frac{\partial \Xi_{ij}}{\partial \rho_{n+1}} \tau_i \tau_j, \quad (\text{A7})$$

with implicit summations over i, j between 1 and $n+1$. The constraint (A7) allows one to determine the value of τ_{n+1} in terms of τ_ν . To do so, let us first express $\partial \Xi_{ij} / \partial \rho_i$ as a function of the derivatives of Σ . By differentiating $\Xi \cdot \Sigma = \mathbb{I}$, we get

$$\frac{\partial \Xi}{\partial \rho_i} = -\Xi \cdot \frac{\partial \Sigma}{\partial \rho_i} \cdot \Xi. \quad (\text{A8})$$

One can also write this relation when the inverse matrix $\tilde{\Xi}$ is defined from the covariance matrix of the cells restricted in \mathcal{A} only, $\tilde{\Sigma}$,

$$\frac{\partial \tilde{\Xi}}{\partial \tilde{\rho}_\kappa} = -\tilde{\Xi} \cdot \frac{\partial \tilde{\Sigma}}{\partial \tilde{\rho}_\kappa} \cdot \tilde{\Xi}. \quad (\text{A9})$$

Substituting $V = \Xi \cdot \tau$ and $\tilde{V} = \tilde{\Xi} \cdot \tilde{\tau}$ into equations (A4), (A6), (A7), we get

$$\lambda_\kappa = \tilde{V}_\kappa \frac{d\tilde{\tau}_\kappa}{d\tilde{\rho}_\kappa} - \frac{1}{2} \tilde{V}^t \cdot \frac{\partial \tilde{\Sigma}}{\partial \tilde{\rho}_\kappa} \cdot \tilde{V}, \quad (\text{A10})$$

$$\lambda_\kappa = V_\kappa \frac{d\tau_\kappa}{d\rho_\kappa} - \frac{1}{2} V^t \cdot \frac{\partial \Sigma}{\partial \rho_\kappa} \cdot V, \quad (\text{A11})$$

$$0 = V_{n+1} \frac{d\tau_{n+1}}{d\rho_{n+1}} - \frac{1}{2} V^t \cdot \frac{\partial \Sigma}{\partial \rho_{n+1}} \cdot V. \quad (\text{A12})$$

An obvious solution to this set of equations is given by

$$V_\mu = \tilde{V}_\mu, V_{n+1} = 0. \quad (\text{A13})$$

Indeed, in that case, using $\partial_i = \partial / \partial \rho_i$ and $\tilde{\partial}_\kappa = \partial / \partial \tilde{\rho}_\kappa$, we have $V_\kappa (d_\kappa \tau_\kappa) = \tilde{V}_\kappa (d_\kappa \tilde{\tau}_\kappa)$, $V^t \cdot \partial_\kappa \Sigma \cdot V = \tilde{V}^t \cdot \tilde{\partial}_\kappa \tilde{\Sigma} \cdot \tilde{V} + 0$ and $V^t \cdot \partial_{n+1} \Sigma \cdot V = 0$ because Σ_{ij} solely depends on the densities ρ_i and ρ_j .

The solution given by equation (A13) is equivalent to

$$\tau_\mu = \tilde{\tau}_\mu, \tau_{n+1} = -\frac{\Xi_{\mu n+1}}{\Xi_{n+1 n+1}} \tilde{\tau}_\mu = \sigma_{n+1 \mu} \tilde{\Xi}_{\mu\nu} \tilde{\tau}_\nu. \quad (\text{A14})$$

We then immediately get

$$\lambda_\mu \rho_\mu = \lambda_\mu \tilde{\rho}_\mu + \frac{1}{2} \tau_i V_i = \frac{1}{2} \tilde{\tau}_\nu \tilde{V}_\nu, \quad (\text{A15})$$

so that the property given by equation (A2), namely $\varphi(\{\lambda_1, \dots, \lambda_n\}) = \varphi(\{\lambda_1, \dots, \lambda_n, \lambda_{n+1} = 0\})$, is established. Note that the same argument holds for the bias function so that $b_\varphi(\lambda_1, \dots, \lambda_n) = b_\varphi(\lambda_1, \dots, \lambda_n, 0)$.

APPENDIX B: LARGE DISTANCE LIMIT

Let us present an heuristic demonstration of the property (38). The argument is based on a peak-background type construction. So let us consider a functional $\mathcal{F}_i[\rho(\mathbf{x}); \mathbf{x}_i]$ of the density field $\rho(\mathbf{x})$ in the vicinity of the location \mathbf{x}_i . The idea is that $\mathcal{F}_i[\rho(\mathbf{x}); \mathbf{x}_i]$ depends on $\rho(\mathbf{x})$ only when $|\mathbf{x} - \mathbf{x}_i|$ is small enough, say $|\mathbf{x} - \mathbf{x}_i| \leq \mathcal{R}$ and we are interested in how $\mathcal{F}_i[\rho(\mathbf{x}); \mathbf{x}_i]$ and $\mathcal{F}_j[\rho(\mathbf{x}); \mathbf{x}_j]$ are correlated when $|\mathbf{x}_1 - \mathbf{x}_2| \gg R$. One can then say that the density field values at large separation are correlated only through the large scale fluctuations, or more precisely through the response function

of the functional \mathcal{F} to large scale density fluctuations. We assume the latter to be in the linear regime and we note them $\delta_L(\mathbf{x})$. Note that $\rho(\mathbf{x})$ is a non trivial functional of $\delta_L(\mathbf{x})$ through mode coupling. Formally in the vicinity of \mathbf{x}_1 for instance we could write

$$\rho(\mathbf{x}) = \rho_S(\mathbf{x}) + \mathcal{D}_L[\rho(\mathbf{x}); \mathbf{x}_1] \delta_L(\mathbf{x}_1) + \dots \quad (\text{B1})$$

where $\rho_S(\mathbf{x})$ is the value of the field when long wave modes have been suppressed and where $\mathcal{D}_L[\rho(\mathbf{x}); \mathbf{x}_1]$ gives the linear response function of the field $\rho(\mathbf{x})$ to the large scale fluctuation δ_L in the vicinity of \mathbf{x}_1 . Then any functional $\mathcal{F}[\rho(\mathbf{x}); \mathbf{x}_1]$ of the density field can be similarly expanded as a function of the local density contrast

$$\mathcal{F}[\rho(\mathbf{x}); \mathbf{x}_1] = \mathcal{F}[\rho_S(\mathbf{x}); \mathbf{x}_1] + \mathcal{D}_L[\mathcal{F}[\rho(\mathbf{x}); \mathbf{x}_1]; \mathbf{x}_1] \delta_L(\mathbf{x}_1) + \dots \quad (\text{B2})$$

Note that in this expression both $\mathcal{F}[\rho_S(\mathbf{x}); \mathbf{x}_1]$ and $\mathcal{D}_L[\mathcal{F}[\rho(\mathbf{x}); \mathbf{x}_1]; \mathbf{x}_1]$ are random quantities that depend on the small scale density fluctuations, independently on the long wave modes collected in $\delta_L(\mathbf{x}_1)$. When taking ensemble average of combination of such quantities, one can observe that

$$\langle \mathcal{F}[\rho(\mathbf{x}); \mathbf{x}_1] \rangle = \langle \mathcal{F}[\rho_S(\mathbf{x}); \mathbf{x}_1] \rangle \quad (\text{B3})$$

at dominant order in δ_L and then that

$$\begin{aligned} \langle \mathcal{F}[\rho(\mathbf{x}); \mathbf{x}_1] \rho(\mathbf{x}_2) \rangle &= \langle \mathcal{F}[\rho_S(\mathbf{x}); \mathbf{x}_1] \rangle \\ &+ \langle \mathcal{D}_L[\mathcal{F}[\rho_S(\mathbf{x}); \mathbf{x}_1]; \mathbf{x}_1] \rangle \xi_L(\mathbf{x}_1, \mathbf{x}_2) \\ &+ \dots \end{aligned} \quad (\text{B4})$$

where

$$\xi_L(\mathbf{x}_1, \mathbf{x}_2) \equiv \langle \delta_L(\mathbf{x}_1) \delta_L(\mathbf{x}_2) \rangle. \quad (\text{B5})$$

The key point is to further observe that,

$$\begin{aligned} \langle \mathcal{F}[\rho(\mathbf{x}); \mathbf{x}_1] \mathcal{F}[\rho(\mathbf{x}); \mathbf{x}_2] \rangle &= \\ \langle \mathcal{F}[\rho_S(\mathbf{x}); \mathbf{x}_1] \rangle \langle \mathcal{F}[\rho_S(\mathbf{x}); \mathbf{x}_2] \rangle &+ \langle \mathcal{D}_L[\mathcal{F}[\rho_S(\mathbf{x}); \mathbf{x}_1]; \mathbf{x}_1] \rangle \langle \mathcal{D}_L[\mathcal{F}[\rho_S(\mathbf{x}); \mathbf{x}_2]; \mathbf{x}_2] \rangle \xi_L(\mathbf{x}_1, \mathbf{x}_2) \\ &+ \dots \end{aligned} \quad (\text{B6})$$

When taking the connected part of these moments, these forms lead to the functional relation,

$$\begin{aligned} \langle \mathcal{F}[\rho(\mathbf{x}); \mathbf{x}_1] \mathcal{F}[\rho(\mathbf{x}); \mathbf{x}_2] \rangle_c &= \frac{1}{\xi_L(\mathbf{x}_1, \mathbf{x}_2)} \\ &\times \langle \mathcal{F}[\rho(\mathbf{x}); \mathbf{x}_1] \rho(\mathbf{x}_2) \rangle_c \langle \mathcal{F}[\rho(\mathbf{x}); \mathbf{x}_2] \rho(\mathbf{x}_1) \rangle_c \end{aligned} \quad (\text{B7})$$

Note that it can easily be extended to two different functionals of the density field and when applied to products of density fields, it precisely leads to the relation (38). This construction actually reproduces the expansion scheme developed in Bernardeau et al. (2008) in Fourier space. In this context the $\langle \mathcal{D}_L[\mathcal{F}[\rho_S(\mathbf{x}); \mathbf{x}_1]; \mathbf{x}_1] \rangle$ is simply the nonlinear propagator in real space. The relation (B6) could be extended to higher order contributions provided one knew how to compute multi-point propagators.

APPENDIX C: COSMIC ERROR STATISTICS

In this section, we show how to go from equation (5) to the cosmic variance of discrete counts. Hence let us assume that

the joint PDF of the density in N_t cells reads

$$\mathcal{P}(\{\hat{\rho}_1\}, \dots, \{\hat{\rho}_{N_t}\}) = \prod_{I=1}^{N_t} \mathcal{P}(\hat{\rho}_I) \left[1 + \sum_{i < j} b(\hat{\rho}_I) b(\hat{\rho}_J) \xi_{IJ} \right]. \quad (\text{C1})$$

In what follows, we will take $\xi_{IJ} = \xi$ independent from the positions and we will denote the density bias $b_I = b(\hat{\rho}_I)$.

C1 The one-cell count-in-cell PDF

We want to estimate $\mathcal{P}(\hat{\rho}) d\hat{\rho}$ by measuring the density in N_t spheres of same radius R . If the spheres are independent i.e there is no bias in the measure $b_I = 0$, then the probability of having N spheres with density $\hat{\rho} \pm \Delta\rho/2$ is simply given by a binomial distribution

$$P^{\text{unbiased}}(N) = \binom{N_t}{N} p^N (1-p)^{N_t-N}, \quad (\text{C2})$$

with $\binom{N_t}{N} = N_t! / N! (N_t - N)!$ and p , the probability for the density of one sphere to be in $\hat{\Delta} = [\hat{\rho} - \Delta\rho/2, \hat{\rho} + \Delta\rho/2]$, is given by

$$p = \int_{\hat{\Delta}} d\rho \mathcal{P}(\rho). \quad (\text{C3})$$

In the Poisson limit (where $N_t \gg N$ and $p N_t$ is constant), using the Stirling formula $x! \approx \sqrt{2\pi x} (x/e)^x$ when x goes to infinity, this probability distribution trivially becomes a Poisson distribution parametrized by $\bar{N} = p N_t$

$$P^{\text{unbiased}}(N) \approx \frac{\bar{N}^N}{N!} \exp(-\bar{N}). \quad (\text{C4})$$

Accounting for a non-zero bias in equation C1, the probability of having N spheres with density in $\hat{\Delta} = \hat{\rho} \pm \Delta\rho/2$ becomes

$$P^{\text{biased}}(N) = \binom{N_t}{N} P\left(\begin{array}{c} \hat{\rho}_1, \dots, \hat{\rho}_N \in \hat{\Delta} \\ \hat{\rho}_{N+1}, \dots, \hat{\rho}_{N_t} \notin \hat{\Delta} \end{array}\right), \quad (\text{C5})$$

where the probability of having only the first N densities in $\hat{\Delta}$ reads

$$\begin{aligned} P\left(\begin{array}{c} \hat{\rho}_1, \dots, \hat{\rho}_N \in \hat{\Delta} \\ \hat{\rho}_{N+1}, \dots, \hat{\rho}_{N_t} \notin \hat{\Delta} \end{array}\right) &= \\ \left(\int_{\hat{\Delta}} d\hat{\rho}_1 \dots d\hat{\rho}_N \left(\int_{\mathbb{R}_+/\hat{\Delta}} d\hat{\rho}_{N+1} \dots d\hat{\rho}_{N_t} \mathcal{P}(\hat{\rho}_1, \dots, \hat{\rho}_{N_t})\right)\right) & \end{aligned}$$

Given equation (C1), equation (C5) eventually yields

$$\begin{aligned} P^{\text{biased}}(N) &= \binom{N_t}{N} \times \left[p^N (1-p)^{N_t-N} \right. \\ &+ \xi \frac{N(N-1)}{2} (pb)^2 p^{N-2} (1-p)^{N_t-N} \\ &+ \xi N(N_t - N) p^{N-1} pb(-pb)(1-p)^{N_t-N-1} \\ &\left. + \xi \frac{(N_t - N)(N_t - N - 1)}{2} p^N (-pb)^2 (1-p)^{N_t-N-2} \right], \end{aligned}$$

where we introduced pb such that

$$pb = \int_{\hat{\Delta}} d\hat{\rho} \mathcal{P}(\hat{\rho}) b(\hat{\rho}) \quad (\text{C6})$$

that comes with complement $-pb$ as the normalisation of \mathcal{P} enforces $\int_{\mathbb{R}_+} d\hat{\rho} \mathcal{P}(\hat{\rho}) b(\hat{\rho}) = 0$. In full generality, note that

the factors ξ entering in the expression of the biased count probability should be understood respectively (and in the order of appearance) as the mean i) autocorrelation between spheres of density in $\hat{\delta}$; ii) cross-correlation between spheres of that density and the others; iii) auto-correlation between the spheres of density not in $\hat{\delta}$. For the sake of simplicity, we will consider here that it is given by the mean correlation $\xi = 2 \sum_{I < J} \xi_{IJ} / [N_t(N_t - 1)]$.

The Poisson limit is then simply given by

$$P^{\text{biased}}(N) \approx \frac{\bar{N}^N}{N!} e^{-\bar{N}} \left[1 + b^2 \xi \left(\frac{N(N-1)}{2} - N\bar{N} + \frac{\bar{N}^2}{2} \right) \right],$$

where $b = pb/p$. It is then straightforward to check that the probability distribution P^{biased} i) is normalised; ii) has mean $\langle N \rangle = \bar{N}$ (and is therefore unbiased); iii) has variance $\langle N^2 \rangle - \langle N \rangle^2 = \bar{N} + b^2 \xi \bar{N}^2$. For large enough \bar{N} , sampling errors can be neglected and the cosmic variance is directly proportional to b^2 .

C2 Cross-correlations

Let us now show how the estimate of the one-cell PDF in two distinct bins are correlated. If the spheres are independent (i.e the joint PDF is unbiased), then the probability, $P^{\text{unbiased}}(N_1, N_2)$, of getting N_1 spheres with density in $\hat{\Delta}_1$ and N_2 spheres with density in $\hat{\Delta}_2$ is given by

$$P^{\text{unbiased}} = \binom{N_t}{N_1} \binom{N_t - N_1}{N_2} p_1^{N_1} p_2^{N_2} (1 - p_1 - p_2)^{N_t - N_1 - N_2},$$

where p_1 is the probability for the density of one sphere to be in $\hat{\Delta}_1$ and p_2 to be in $\hat{\Delta}_2$. In the Poisson limit, this probability distribution trivially becomes the product of two Poisson distribution parametrized by $\bar{N}_1 = p_1 N_t$ and $\bar{N}_2 = p_2 N_t$

$$P^{\text{unbiased}}(N_1, N_2) \approx P^{\text{unbiased}}(N_1) P^{\text{unbiased}}(N_2). \quad (\text{C7})$$

As expected, in the case where there is no spatial correlation between spheres, the estimate of the density PDF in each bin is independent.

In the biased case, the probability of having N_1 spheres with density in $\hat{\Delta}_1$ and N_2 spheres with density in $\hat{\Delta}_2$ becomes

$$P^{\text{biased}}(N_1, N_2) = P^{\text{unbiased}}(N_1, N_2) [1 + \xi \mathfrak{b}], \quad (\text{C8})$$

where \mathfrak{b} is given by

$$\begin{aligned} \mathfrak{b} = & \frac{N_1(N_1 - 1)}{2} b_1^2 + N_1 N_2 b_1 b_2 + \frac{N_2(N_2 - 1)}{2} b_2^2 \\ & + \frac{(N_t - N_1 - N_2)(N_t - N_1 - N_2 - 1)}{2} \left(\frac{b_1 p_1 + b_2 p_2}{1 - p_1 - p_2} \right)^2 \\ & + (N_1 b_1 + N_2 b_2) N_t \frac{b_1 p_1 + b_2 p_2}{1 - p_1 - p_2}, \end{aligned}$$

the bias parameters b_1 and b_2 being defined as

$$b_i = \frac{\int_{\hat{\Delta}_i} d\hat{\rho} \mathcal{P}(\hat{\rho}) b(\hat{\rho})}{\int_{\hat{\Delta}_i} d\hat{\rho} \mathcal{P}(\hat{\rho})}. \quad (\text{C9})$$

In the Poisson limit, equation (C8) becomes

$$\begin{aligned} P^{\text{biased}}(N_1, N_2) \approx P^{\text{unbiased}}(N_1, N_2) \left[1 \right. \\ & + \xi b_1^2 \left(\frac{N_1(N_1 - 1)}{2} + \frac{\bar{N}_1^2}{2} - N_1 \bar{N}_1 \right) \\ & + \xi b_2^2 \left(\frac{N_2(N_2 - 1)}{2} + \frac{\bar{N}_2^2}{2} - N_2 \bar{N}_2 \right) \\ & \left. + \xi b_1 b_2 (N_1 N_2 + \bar{N}_1 \bar{N}_2 - N_1 \bar{N}_2 - \bar{N}_1 N_2) \right]. \end{aligned}$$

This probability distribution $P^{\text{biased}}(N_1, N_2)$ is normalised, its marginals are respectively $P^{\text{biased}}(N_1)$ and $P^{\text{biased}}(N_2)$ and the covariance between N_1 and N_2 is given by

$$\langle N_1 N_2 \rangle = \bar{N}_1 \bar{N}_2 (1 + \xi b_1 b_2). \quad (\text{C10})$$

It should be straightforward, if tedious, to generalize equation (C8) to an arbitrary number of cells.

APPENDIX D: CONSISTENCY OF BIAS

Let us measure $b^2(\hat{\rho})$ from two estimators and check for consistency. The first is based on cross-correlations and used in Section 6.1 to measure the density bias function (see equation (48)). We use the following estimator to measure $b^2(\hat{\rho})$ in this context,

$$\hat{b}_{\text{cross}}^2(\hat{\rho}) = \frac{1}{\xi^2} \left[\frac{\sum_I \sum_{j=1}^6 \epsilon(\rho_I, \hat{\rho}, \hat{\Delta}) \rho_{\alpha_{I,j}}}{6 \sum_I \epsilon(\rho_I, \hat{\rho}, \hat{\Delta})} - 1 \right]^2, \quad (\text{D1})$$

where the sum is over each sphere $1 \leq I \leq N_t$ of radius $R = 10 \text{Mpc}/h$ with density ρ_I and its 6 neighbours at distance $r_e = 20 \text{Mpc}/h$ labelled with the indices $\alpha_{I,j}$ for $1 \leq j \leq 6$, $\epsilon(x, \hat{\rho}, \hat{\Delta}) = \mathcal{B}(x \in [\hat{\rho} - \Delta\rho/2, \hat{\rho} + \Delta\rho/2])$ with \mathcal{B} a boolean function which evaluates to one if the density is in a bin centred on $\hat{\rho}$ with width $\Delta\rho = 3/21$ and the measured dark matter correlation function at distance r_e is given by equation (66).

Alternatively, one can use equation (47) to estimate the density bias squared via the auto-correlation of cells

$$\hat{b}_{\text{auto}}^2(\hat{\rho}) = \frac{1}{\xi} \left[\frac{N_t \sum_{I,j} \epsilon(\rho_I, \hat{\rho}, \hat{\Delta}) \epsilon(\rho_{\alpha_{I,j}}, \hat{\rho}, \hat{\Delta})}{6 \sum_I \epsilon(\rho_I, \hat{\rho}, \hat{\Delta}) \sum_{I,j} \epsilon(\rho_{\alpha_{I,j}}, \hat{\rho}, \hat{\Delta})} - 1 \right]. \quad (\text{D2})$$

Fig. D1 compares the mean and error on the mean of the density bias estimated by equations (D1) and (D2) when the simulation is divided into eight sub-cubes. Both approaches seem to give consistent result in particular for intermediate densities. Note that in the case of the auto-correlation (equation (D2)), the statistics is expected to be poorer and the estimator therefore noisier. Indeed, the frequency of occurrence of the event “the two densities are in the same bin” is low in particular in the tails of the distribution and it is very likely that our error bars are under-estimated in this case.

APPENDIX E: ASYMPTOTIC BEHAVIOUR

The asymptotic behaviour of the density bias $b(\hat{\rho})$ can be predicted by integrating equations (41) and (42) around the critical point which dominates the complex integration for $\hat{\rho}$ sufficiently large compared to ρ_c . Bernardeau et al. (2015)

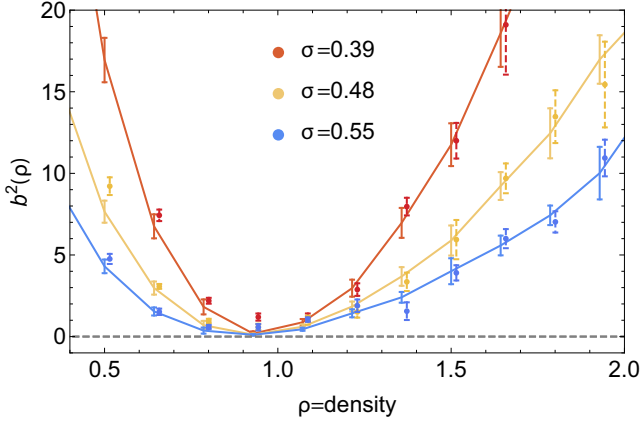


Figure D1. Density bias squared measured in the simulation from the auto-correlation (dashed error bars) and cross-correlation (solid line and error bars) estimators given by equations (D1) and (D2) at different variances as labelled. Both approaches give consistent results.

indeed showed that in this regime, the cumulant generating function behaves like

$$\varphi(\lambda) = \varphi_c + (\lambda - \lambda_c)\rho_c + (\lambda - \lambda_c)^{3/2} \frac{2\sqrt{2}}{3\sqrt{\pi_3}} - (\lambda - \lambda_c)^2 \frac{\pi_4}{6\pi_3^2} + \dots$$

with $\pi_i = \Psi^{(i)}(\rho_c)$, the successive derivatives of the rate function at the critical point. This expansion for the cumulant generating function can be obtained by first inverting, around the critical point, the relation between ρ and λ coming from equation (20)

$$\lambda = \lambda_c + \sum_i \frac{(\rho - \rho_c)^i}{i!} \pi_{i+1}, \quad (\text{E1})$$

which reads

$$\rho = \rho_c + \sqrt{\frac{2(\lambda - \lambda_c)}{\pi_3}} - \frac{\pi_4(\lambda - \lambda_c)}{3\pi_3^2} + \dots,$$

and plugging the result into equation (19).

The large-density tail of the density PDF and density bias can then be derived by computing the inverse Laplace transform of the generating function $\varphi(\lambda)$ when the integrand of equations (41) and (42) is dominated by its singular part, near $\lambda \approx \lambda_c$. In that case, Balian & Schaeffer (1989) and BPC showed that the dominant contribution of the path in the complex plane is along the real axis and wrapping around the singular value λ_c as illustrated on Fig. E1. As the two branches $\mathbb{R}_+ + \lambda_c \pm i\epsilon$ of that path yield complex conjugate contributions, the density PDF can be approximated by

$$P(\hat{\rho}) \approx \Im \left\{ \int_{i\epsilon + \lambda_c}^{i\epsilon + \infty} \frac{d\lambda}{\pi} \exp[\varphi_c - \lambda_c \hat{\rho} - (\lambda - \lambda_c)(\hat{\rho} - \rho_c)] \times \left[1 + \frac{2\sqrt{2}}{3\sqrt{\pi_3}} (\lambda - \lambda_c)^{3/2} + \dots \right] \right\}, \quad (\text{E2})$$

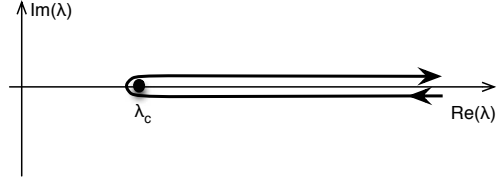


Figure E1. The path line in the λ complex plane used for the computation of the large-density asymptotic forms follows a so-called Hankel contour.

and the density bias by

$$b(\hat{\rho})P(\hat{\rho}) \approx \Im \left\{ \int_{i\epsilon + \lambda_c}^{i\epsilon + \infty} \frac{d\lambda}{\pi} \exp[\varphi_c - \lambda_c \hat{\rho} - (\lambda - \lambda_c)(\hat{\rho} - \rho_c)] \times \left[1 + \frac{2\sqrt{2}}{3\sqrt{\pi_3}} (\lambda - \lambda_c)^{3/2} + \dots \right] b_\varphi(\lambda) \right\}, \quad (\text{E3})$$

where \Im is the imaginary part.

As already shown in Bernardeau et al. (2015), the resulting large-density PDF can then be written, to leading order, as

$$\mathcal{P}(\hat{\rho} \gg \rho_c) = \frac{1}{\pi} \exp(\varphi_c - \lambda_c \hat{\rho}) \Im \left[\frac{\sqrt{\pi}}{\sqrt{2\pi_3}} (\hat{\rho} - \rho_c)^{-5/2} + \dots \right],$$

while the corresponding bias function is given to leading order as the ratio of equations (E3) and (E2)

$$b(\hat{\rho} \gg \rho_c) = (\hat{\rho} - \rho_c) V^{(1)}(\rho_c) + V(\rho_c) - \frac{\pi_4}{2\pi_3^2} V^{(2)}(\rho_c) + \frac{1}{2\pi_3} V^{(3)}(\rho_c) + \mathcal{O}\left(\frac{1}{\hat{\rho} - \rho_c}\right) \quad (\text{E4})$$

where in the one-cell case $V(\rho) = \Xi_{ij} \tau_j(\rho_j)$ is given by $V(\rho) = \zeta^{-1}(\rho)/\sigma^2(R\rho^{1/3}, R\rho^{1/3})$ and is nothing but the bias function $V(\rho) = b_\varphi(\lambda = \Psi'(\rho))$. In particular, one can compute the successive derivatives of $V(\rho)$ at the critical density, for instance

$$V^{(1)}(\rho_c) = \frac{(\lambda_c \rho_c (\rho_c^{1/\nu} - 1 - 1/\nu) + \frac{1}{\nu} \varphi_c)}{\nu (\rho_c^{1/\nu} - 1)^2 \rho_c^{1-1/\nu}}. \quad (\text{E5})$$

APPENDIX F: APPROXIMATION SCHEMES FOR THE SLOPE

In the context of large-deviation theory, one can compute the rate function of the slope $s = (\rho_2 - \rho_1)R_1/\Delta R$ by applying the so-called contraction principle which yields

$$\Psi_s(\hat{s}) = \inf_{\rho} \Psi\left(\rho, \rho + s \frac{\Delta R}{R_1}\right). \quad (\text{F1})$$

The cumulant generating function of the slope can then be obtained through the Gartner-Ellis theorem as the Legendre transform of the rate function

$$\varphi_s(\mu) = \mu s - \Psi_s(s) \quad \text{with} \quad \mu = \Psi'_s(s). \quad (\text{F2})$$

The slope PDF then follows by Inverse Laplace transform as shown in equation (52). It is now easy to see that the approximations of the slope PDF and bias can be obtained similarly to what has been done for the density. In this case, the slope cumulant generating function has typically two critical points $s_c^- < 0$ and $s_c^+ > 0$ corresponding to the zeros

of the second derivative of the rate function $d^2\Psi_s(s)/ds^2 = 0$.

For small slope $s_c^- < s < s_c^+$, the slope PDF and bias can be approximated by a saddle point approximation of equations (51) and (52) so that

$$\mathcal{P}(\hat{s}) \approx \sqrt{\frac{\Psi_s''(\hat{s})}{2\pi}} \exp(-\Psi_s(\hat{s})), \quad (\text{F3})$$

and

$$b(\hat{s}) \approx b_\varphi(-R_1\Psi'_s(s)/\Delta R, R_1\Psi'_s(s)/\Delta R). \quad (\text{F4})$$

However, this approximation is not sufficient (in particular it would predict $b(0) = 0$ which is far from the true value); one should Taylor expand the bias function b_s around $\mu = \Psi'_s(\hat{s})$ in order to get a better fit to the full numerical integration. This issue is left for future work.

Conversely, the asymptotes of the PDF and bias at large (positive and negative) slopes can be obtained similarly to what was done for the large-density regime (see appendix E). For instance, for large positive slope, the PDF can be approximated by

$$\begin{aligned} \mathcal{P}(\hat{s} \gg s_c^+) &= \frac{1}{\pi} \exp(\varphi_s(\mu_c^+) - \mu_c^+ \hat{s}) \\ &\times \Im \left[\sqrt{\frac{\pi}{2\pi_3^{s,+}}} (\hat{s} - s_c^+)^{-5/2} \right] + \mathcal{O}((\hat{s} - s_c^+)^{-7/2}), \end{aligned}$$

with $\pi_i^{s,+} = \Psi_s^{(i)}(s_c^+)$, while the corresponding slope bias is

$$\begin{aligned} b(\hat{s} \gg s_c^+) &= (\hat{s} - s_c^+) V_s^{(1)}(s_c^+) + V_s(s_c^+) - \frac{\pi_4^s}{2(\pi_3^s)^2} V_s^{(2)}(s_c^+) \\ &+ \frac{1}{2\pi_3^s} V_s^{(3)}(s_c^+) + \mathcal{O}\left(\frac{1}{\hat{s} - s_c^+}\right), \end{aligned}$$

where we define $V_s(s) = b_\varphi(-R_1\Psi'_s(s)/\Delta R, R_1\Psi'_s(s)/\Delta R)$.